

BIBLIOTECA UCM



5301691857

T 524.73
MAM
Bio

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Biológicas

Departamento de Bioquímica y Biología Molecular

BIOLOGÍA MOLECULAR DEL BACTERIÓFAGO Cp-1 DE *Streptococcus pneumoniae*

Memoria que para optar al Grado de Doctor en Ciencias Biológicas presenta:

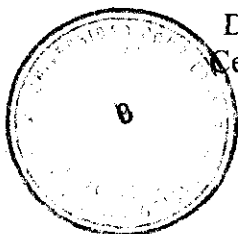
ANA MARÍA DEL CARMEN MARTÍN RODRÍGUEZ

Director de la Tesis

Dr. Pedro García González

Colaborador Científico

Departamento de Microbiología Molecular
Centro de Investigaciones Biológicas (CSIC)



Madrid, 1996

R. 24.673

A mis padres

A Julio

Quiero expresar mi más sincero agradecimiento a todas aquellas personas que, de alguna u otra manera, me han ayudado en la realización de esta Tesis:

De manera especial, quiero agradecer a mi director de Tesis, el Dr. Pedro García, la paciencia que siempre tuvo conmigo, su dedicación y la formación científica que de él he recibido. Nunca olvidaré mis primeros pasos en este mundo de la investigación a su lado.

Por otra parte, deseo expresar mi más sincero agradecimiento al Profesor Rubens López, no sólo por haberme aceptado en su laboratorio sino por sus sabios consejos, comentarios y sugerencias y por su estímulo y cariño en todo momento. Escucharle es todo un placer. Destacar especialmente al Profesor Ernesto García, por su disponibilidad desinteresada en todo momento y por su admirable calidad profesional y humana que demuestra día a día. Resaltar también al Dr. José Luis García por sus innumerables “variantes” propuestas a lo largo de este trabajo. Quisiera también agradecerles su colaboración en la corrección de esta Memoria. Y no quiero olvidar a la Dra. Concepción Ronda, que durante el tiempo que trabajé con ella siempre demostró una enorme simpatía y calidad humana.

Deseo expresar mi más sincero agradecimiento a todos mis compañeros de laboratorio, los que estuvieron pero ya no están y los que están, y cuya larga lista os obviaré, con los que he compartido momentos que no olvidaré, además del trabajo diario.

También quisiera agradecer al laboratorio de la Dra. Margarita Salas, su acogida para realizar una parte de este trabajo, y muy especialmente a Juan, Belén con los que tuve la suerte de poder trabajar y a Luis que nos puso sobre la pista de las repeticiones de los extremos.

Asimismo, agradecer la excelente asistencia técnica de Eloisa Cano, Manuel Carrasco, Paquita Morante, Aurelio Hurtado, Javier Varela, Ricardo Uña, Monika Fontenla y Victoria Muñoz que han facilitado la realización de este trabajo.

Al Dr. Alvaro Martínez por su labor de tutoría en la Facultad de Químicas.

No quiero olvidar a mis amigos/as de Pamplona y Valladolid que a pesar de la distancia siempre estuvieron muy cerca de mí y, a Mari Paz, mi prima y amiga, con la que he compartido mi vida de estudiante y predoctoral.

Para finalizar quiero agradecerles a mis padres a quienes debo todo y, a Julio que tuvo la paciencia de aguantarme durante todo este periodo, sufriendo los éxitos y los fracasos de los experimentos.

El presente trabajo ha sido realizado gracias a la concesión de una Beca de Formación de Personal Investigador del Ministerio de Educación y Ciencia.

*“La ciencia será siempre una búsqueda,
jamás un descubrimiento real. Es un
viaje, nunca una llegada.”*

Karl Popper

ABREVIATURAS Y SÍMBOLOS

Ap	:	ampicilina
ATP	:	adenosina 5'-trifosfato
BrEt	:	bromuro de etidio
BSA	:	seroalbúmina bovina
C-	:	carboxilo
cDNA	:	ácido desoxiribonucleico complementario
Ci	:	curio
Cm	:	cloranfenicol
dATP	:	deoxiadenosina 5'-trifosfato
dCTP	:	deoxicitidina 5'-trifosfato
ddCTP	:	dideoxicitidina 5'-trifosfato
ddTTP	:	dideoxitimidina 5'-trifosfato
dGTP	:	deoxiguanosina 5'-trifosfato
dNTP	:	deoxinucleósidos 5'-trifosfato
dNMP	:	deoxinucleósido 5'-monofosfato
DOC	:	desoxicolato sódico
DSB	:	proteínas de unión a banda doble de DNA
dsDNA	:	DNA de doble banda
dTTP	:	deoxitimidina 5'-trifosfato
DTT	:	ditiotreitól
EDTA	:	ácido etilen-diamino-tetraacético
EGTA	:	ácido etilen-glicol bis (β -aminoetileter) N,N,N',N'tetraacético
Ery	:	eritromicina
gp	:	producto del gen
HPUra	:	6- <i>p</i> -(hidroxifenilazo)-uracil
IPTG	:	isopropil- β -D-tiogalactopiranosido
ITR	:	repeticiones terminales invertidas
kb	:	kilobases
kDa	:	kilodaltons
klenow	:	fragmento klenow de la DNA polimerasa I de <i>E. coli</i>
Km	:	resistencia a la kanamicina
LB	:	medio de cultivo de Luria y Bertani
LTA	:	ácido lipoteicoico
3-MB	:	3-metil-benzoato
m.d.i	:	multiplicidad de infección
M_r	:	masa molecular relativa
N-	:	amino
N	:	unidad de nefelometría
nt	:	nucleótidos
DO	:	densidad óptica
ORF	:	marco de lectura abierto
pb	:	par de bases
PCR	:	reacción en cadena de la polimerasa
pI	:	punto isoeléctrico
pRNA	:	RNA piloto o RNA de las precabezas
p/v	:	relación peso/volumen
R	:	resistencia

RBS	:	sitio de unión al ribosoma
RNAsina	:	inhibidor de ribonucleasas
SDS	:	dodecil sulfato sódico
Sm	:	estreptomicina
SSB	:	proteínas de unión a banda simple de DNA
ssDNA	:	DNA de cadena simple
Tc	:	tetraciclina
Tris	:	2-amino-2-hidroximetilpropano-1,3-diol
TP	:	proteína terminal
TP-DNA	:	TP del fago (Cp-1 o ϕ 29), con proteína terminal unida covalentemente a los extremos 5'
TP-dNMP	:	complejo covalente formado por la TP de (Cp-1 o ϕ 29) y un residuo de dNMP
U	:	unidad
UFC	:	unidades formadoras de colonias
UFP	:	unidades formadoras de placas
UV	:	ultravioleta
v/v	:	relación volumen/volumen

ABREVIATURAS PARA AMINOÁCIDOS

Ala	Alanina	A
Arg	Arginina	R
Asn	Asparagina	N
Asp	Ácido aspártico	D
Cys	Cisteína	C
Gln	Glutamina	Q
Glu	Ácido glutámico	E
Gly	Glicina	G
His	Histidina	H
Ile	Isoleucina	I
Leu	Leucina	L
Lys	Lisina	K
Met	Metionina	M
Phe	Fenilalanina	F
Pro	Prolina	P
Ser	Serina	S
Thr	Treonina	T
Trp	Triptófano	W
Tyr	Tirosina	Y
Val	Valina	V

GLOSARIO DE TÉRMINOS NO CASTELLANOS

A lo largo del texto se ha tratado de evitar anglicismos en la medida de lo posible.. Sólo en los casos que se presentan en la siguiente lista se ha preferido mantener la nomenclatura anglosajona:

“Quellung”. Visibilizar las células de neumococo en una de las técnicas de diagnóstico.

“Primer extension”. Técnica que detecta los puntos exactos de iniciación de la transcripción, utilizando un oligonucleótido como cebador de la reacción.

“Random primer”. Técnica de marcaje radiactivo de sondas de hibridación mediante la presencia de la mezcla de hexanucleótidos de secuencias aleatorias.

“Protein-priming”. Término para describir el mecanismo de iniciación de la replicación que utiliza una proteína como donadora del grupo 3'-OH necesario.

“Sliding-back”. Traslocación del complejo TP-dAMP tras la reacción de iniciación de la replicación. Podría traducirse como retroceso o deslizamiento hacia atrás.

“Jumping-back”. Término para describir un salto hacia atrás tras la reacción de iniciación de la replicación.

ÍNDICE

I- INTRODUCCIÓN	.1
1.GENERALIDADES DE NEUMOCOCO	.2
1.1. Importancia de neumococo desde el punto de vista clínico	.2
1.2. Importancia de neumococo en el desarrollo de la biología	.3
2. BACTERIÓFAGOS DE NEUMOCOCO	.3
2.1. Familias de bacteriófagos que infectan neumococo	.4
2.2. El bacteriófago Cp-1 de <i>S. pneumoniae</i>	.6
2.2.1. Propiedades biológicas.	.8
2.2.2. Propiedades del DNA y de las proteínas del virión	.8
3. REPLICACIÓN DEL DNA.	.9
3.1. Mecanismo de iniciación con proteína o “protein-priming”	.10
3.1.1. Genomas con TP	.11
3.1.2. “Sliding-back” : un nuevo mecanismo para la transición entre la iniciación y la elongación en los genomas con proteína terminal.	.13
4. MECANISMOS DE LISIS: LIBERACIÓN DE LA PROGENIE FÁGICA	.15
4.1. Papel de las holinas y las enzimas líticas en el proceso de lisis	.16
4.1.1. Holinas	.16
4.1.2. Enzimas líticas	.21
4.2. Enzimas líticas de <i>S. pneumoniae</i> y sus bacteriófagos	.25
4.2.1. Autolisinas de neumococo.	.26
4.2.2. Enzimas líticas de los bacteriófagos de <i>S. pneumoniae</i>	.27
5. OBJETIVOS	.31

II- MATERIALES Y MÉTODOS33
1. CEPAS BACTERIANAS, BACTERIÓFAGOS Y PLÁSMIDOS.34
2. TITULACIÓN, CRECIMIENTO Y PURIFICACIÓN DE FAGOS35
3. PREPARACIÓN DEL DNA FÁGICO36
4. MEDIOS Y CONDICIONES DE CULTIVO36
4.1. <i>E. coli</i>36
4.2. <i>S. pneumoniae</i>36
5. PROCEDIMIENTOS DE TRANSFORMACIÓN GENÉTICA.37
5.1. <i>E. coli</i>37
5.2. <i>S. pneumoniae</i>37
6. ELECTROFORESIS DE DNA EN GELES DE AGAROSA38
7. AISLAMIENTO Y PURIFICACIÓN DE FRAGMENTOS DE DNA38
7.1. Técnica de Gene-clean38
7.2. Técnica de la β -agarasa38
7.3. Geles de agarosa de bajo punto de fusión.39
8. TÉCNICA DE LA REACCIÓN EN CADENA DE LA POLIMERASA (PCR) .39	
9. MANIPULACIÓN DEL DNA CON ENZIMAS DE USO COMÚN EN BIOLOGÍA MOLECULAR40
10. TÉCNICAS DE HIBRIDACIÓN DEL DNA Y DEL RNA41
10.1. Técnica de "Southern-blot".41
10.2. Hibridación de DNA sobre filtros de colonias celulares.41
10.3. Técnica de "Dot-blot"42
10.4. Técnica de "Northern-blot".42
11. PREPARACIÓN DE PLÁSMIDOS42
12. SECUENCIACIÓN DE DNA43
12.1. Análisis de la secuencia de DNA43
13. AISLAMIENTO DE RNA43
14. DETERMINACIÓN DE LOS SITIOS DE INICIACIÓN DE LA TRANSCRIPCIÓN MEDIANTE EL EMPLEO DE LA TÉCNICA DE "PRIMER EXTENSION"44

15. ELECTROFORESIS DE LAS PROTEÍNAS EN GELES DE POLIACRILAMIDA-SDS45
16. OBTENCIÓN DE EXTRACTOS DE CULTIVOS CELULARES46
16.1 Extractos preparados de cepas de <i>E. coli</i> y <i>S. pneumoniae</i>46
16.2 Extractos de neumococo infectados con Cp-1 para realizar ensayos de iniciación de replicación48
17. DETECCIÓN DE PROTEÍNAS CODIFICADAS POR FRAGMENTOS CLONADOS DE DNA49
17.1 Técnica de “maxicélulas”49
17.2 Técnica de transcripción-traducción <i>in vitro</i>49
18. PURIFICACIÓN DE PROTEÍNAS POR ELECTROELUCIÓN49
19. OBTENCIÓN DE ANTICUERPOS ANTI-gp9* DEL VIRIÓN49
20. TÉCNICA DE “WESTERN-BLOT”50
21. DETERMINACIÓN DE LA SECUENCIA N-TERMINAL DE PROTEÍNAS .	.51
22. DETERMINACIÓN DEL FENOTIPO Lyt DE NEUMOCOCO51
23. ENSAYOS DE ACTIVIDAD LÍTICA51
24. DETECCIÓN DE HEMOLISINA52
25. REPLICACIÓN DE DNA <i>IN VITRO</i>52
25.1. Formación <i>in vitro</i> del complejo de iniciación TP-dNMP52
25.2. Ensayos de elongación truncada.53
 III- RESULTADOS54
1. SECUENCIA DE NUCLEÓTIDOS DEL GENOMA DE Cp-155
1.1. Estrategia de secuenciación55
1.2. Análisis de la secuencia de nucleótidos58
1.3. Comparación de la secuencia de aminoácidos72

2. MAPA DE TRANSCRIPCIÓN DEL BACTERIOFAGO Cp-1	79
2.1. Determinación del sitio de iniciación de la transcripción.	80
2.2. Mapa de transcripción de los genes tempranos	87
2.3. Mapa de transcripción de los genes tardíos	90
2.4. RNA implicado en la encapsidación del DNA	93
3. ANÁLISIS DE LAS ORFs DEL DNA DE Cp-1	95
3.1. Clonación y expresión de la ORF4 que codifica la TP y de la ORF5 que codifica la DNA polimerasa	95
3.2. Determinación de la secuencia N-terminal de las proteínas codificadas por las ORFs 1, 2, 6, 7, 10, 15, 16 y 17 en <i>E. coli</i>	99
3.3. Estudios sobre la proteína mayoritaria del virión, producto de la ORF9	100
3.3.1. Procesamiento de la proteína mayoritaria del virión	100
3.3.2. La proteasa que codifica la proteína mayoritaria del virión está codificada por el propio fago	100
3.4. Análisis de las ORFs implicadas en los procesos de lisis	109
3.4.1. Clonación de la ORF21 que codifica la holina y de la ORF22 que codifica la lisozima en <i>E. coli</i>	109
3.4.2. Expresión inducible de los genes <i>cph1</i> y <i>cpl1</i> en <i>E. coli</i>	111
3.4.3. El producto del gen <i>cph1</i> complementa la mutación <i>Sam7</i> del fago lambda	111
3.4.4. Inducción de la actividad de LytA en <i>E. coli</i>	113
3.4.5. Clonación de los genes <i>cph1</i> y <i>cpl1</i> en neumococo	116
3.4.6. Expresión de los genes <i>cph1</i> y <i>cpl1</i> en neumococo	116
4. INICIACIÓN DE LA REPLICACIÓN DEL DNA DE CP-1.	121
4.1. Formación del complejo de iniciación usando oligonucleótidos (29-mer) con la secuencia del extremo del genoma del fago Cp-1	122
4.2. La iniciación de la replicación del DNA de Cp-1 está dirigido por el tercer nucleótido	124
4.3. Especificidad del origen de replicación: sitio de iniciación de extractos de Cp-1 sobre secuencias terminales del DNA de $\phi 29$	128
4.4. Replicación de los oligonucleótidos de ssDNA. Recuperación de la primera y segunda T	129
IV- DISCUSIÓN	134
1. ORGANIZACIÓN FUNCIONAL DEL GENOMA DE Cp-1	135
2. ANÁLISIS DE PROTEÍNAS CODIFICADAS POR Cp-1	143
3. MECANISMO DE INICIACIÓN DE LA REPLICACIÓN	154

V- CONCLUSIONES	160
----------------------------------	------------

VI- BIBLIOGRAFÍA	163
-----------------------------------	------------

I. INTRODUCCIÓN

1. GENERALIDADES DE NEUMOCOCO

1.1. Importancia de neumococo desde el punto de vista clínico

Streptococcus pneumoniae (neumococo) es una bacteria Gram-positiva, de 0,5-1,25 μm de diámetro, de forma oval con la extremidad distal lanceolada, inmóvil, y que no forma endosporas. Esta bacteria se presenta en forma diploidea, de ahí el nombre de *Diplococcus pneumoniae* con el que se le denominó al ser descubierto. Neumococo se aisló por primera vez en 1881 por Sternberg, en EE.UU., y Pasteur, en Francia, a partir de muestras de saliva humana (revisado por Austrian, 1981).

Un año después de su aislamiento se identificó este microorganismo como el agente causante de la neumonía, en un trabajo en el que se describe, por primera vez, la tinción de Gram. Neumococo es un patógeno humano responsable no sólo de la neumonía, sino de enfermedades como meningitis y bacteriemias, y con altas tasas de mortalidad y morbilidad. Según la Organización Mundial de la Salud (OMS), la neumonía es actualmente la cuarta causa de mortalidad en el mundo y neumococo, en concreto, es el agente causal de entre el 10 y el 25% de los casos (Sahn, 1990). Además de infecciones invasivas graves, puede producir otitis media y sinusitis. Coloniza las vías altas del sistema respiratorio y, más concretamente, la nasofaringe desde los primeros días de vida. Se considera, por tanto, que cualquier individuo ha estado colonizado por neumococo en alguna etapa de su vida. La frecuencia de portadores puede cambiar de unas zonas geográficas a otras dependiendo de muchos factores (edad, ambiente, estación del año, etc.) y se ha estimado que más del 60% de la población puede ser portadora (Austrian, 1986).

A comienzos de este siglo Neufeld descubrió la acción lítica de las sales biliares (desoxicolato sódico, DOC) sobre neumococo (Neufeld, 1900, 1902), y la denominada reacción de hinchamiento capsular (“quellung”) en presencia de anticuerpos específicos, que se han convertido en dos de las pruebas de diagnóstico habitualmente utilizadas en clínica para la identificación de neumococo. Sin embargo, cada vez es más frecuente en los análisis clínicos de identificación el uso de sondas de DNA específicas de neumococo (Fenoll y cols., 1990).

En 1912, Morgenroth y Kaufmann describieron en neumococo el primer caso de resistencia *in vivo* frente a una droga antibacteriana, la optoquina (un derivado de la

quinina). Hoy en día la resistencia de esta bacteria a agentes antimicrobianos se ha incrementado notablemente, lo que le ha situado, nuevamente, a neumococo entre las bacterias emergentes por su importancia clínica.

1.2. Importancia de neumococo en el desarrollo de la biología

El estudio de la cápsula y de la pared celular de neumococo ha sido de gran utilidad para el desarrollo de la inmunología, ya que ha permitido el análisis de las defensas celular y humoral contra la infección bacteriana, además ha propiciado la caracterización de polisacáridos como la primera sustancia no proteica que mostraba antigenicidad (la cápsula y el ácido teicoico de la pared celular de neumococo). Asimismo ha sido neumococo la primera bacteria en la que se ha estudiado la naturaleza enzimática de la lisis bacteriana (Dubos, 1937) y donde se ha clonado el gen que codifica una autolisina bacteriana (García y cols., 1985a, 1986a). Sin embargo, de todas las aportaciones a la biología y a la medicina derivadas del estudio de neumococo, la de mayor transcendencia fue la que permitió la elucidación del fenómeno de transformación genética (Griffith, 1928), que culminó con el descubrimiento del DNA como responsable de la información genética. Este descubrimiento ha sido la piedra angular para el nacimiento y desarrollo de la biología molecular (Avery y cols., 1944).

2. BACTERIÓFAGOS DE NEUMOCOCO

Como ya se ha señalado, la transformación genética de los tipos capsulares de neumococo supuso un hito histórico en biología. No obstante, el posterior desarrollo de la biología molecular en este importante patógeno humano se ha visto limitado por diversos problemas técnicos, tales como la exigencia de medios de cultivo complejos para su multiplicación en el laboratorio. Este hecho podría explicar, en gran medida, tanto las dificultades para aislar y caracterizar un número adecuado de mutantes que facilitarían los estudios genéticos, como la carencia de condiciones experimentales idóneas para detectar placas fágicas.

Por otra parte, los virus bacterianos, también llamados bacteriófagos o abreviadamente fagos, desde que fueron descubiertos por Twort (1915) y d'Hérelle (1917),

han contribuido en gran medida al espectacular desarrollo experimentado por la biología molecular, en los últimos años. Los primeros fagos de *S. pneumoniae* se aislaron en 1975 a partir de frotis de garganta obtenidos de individuos sanos, por dos grupos de investigación que trabajaban independientemente, el fago Dp-1 por McDonnell y cols. (1975) y el fago ω por Tiraby y cols. (1975). Desde un principio, estos fagos proporcionaron una importante herramienta de trabajo para confirmar y ampliar los estudios de transformación genética la cual, hasta entonces, representaba el único medio de analizar, en este sistema, la transferencia del DNA y la recombinación. No obstante, en los últimos veinte años se han aislado fagos de neumococo en varios países (Bernheimer y cols., 1979; Ronda y cols., 1981; López y cols., 1977, 1981, 1982; Díaz y cols., 1992a) y se ha puesto de manifiesto la gran variabilidad que existe entre los mismos, ya que se han identificado fagos líticos y atemperados con morfologías muy diversas, y se han analizado una serie de interesantes peculiaridades que les han convertido en modelos muy adecuados de estudios a nivel molecular en el sistema de neumococo.

2.1. Familias de bacteriófagos que infectan a neumococo

Los primeros fagos líticos aislados, como hemos dicho anteriormente se denominaron Dp-1 y ω , en 1975. y los primeros fagos atemperados, aislados a partir de estirpes clínicas y capsuladas de neumococo por H. Bernheimer (1977, 1979), se designaron como HB-1, HB-2, HB-3, HB-623, HB-746, entre otros. Esta investigadora sugirió además una interrelación entre lisogenia y patogenicidad (Bernheimer, 1979).

Los fagos de neumococo analizados pertenecen a cuatro familias y hasta el momento se han identificado varios fagos líticos (Dp-1, ω , Cp-1) y atemperados (HB-3 y EJ-1) con características serológicas distintas y morfologías tan diversas como las que se observan en la Figura 1. Dp-1 y HB-3 pertenecen a la familia *Siphoviridae*, con cola larga no contráctil, Cp-1 a la *Podoviridae*, con cola corta y EJ-1 a la *Myoviridae*. El bacteriófago EJ-1, no se incluía en las clasificaciones de las colecciones de virus tipo (ICTV) porque no se ha conseguido propagar *in vitro*, a no ser que se induzca la cepa lisogénica con mitomicina C, pero recientemente Ackermann y cols. (1995) le han clasificado dentro del grupo A1 de la familia *Myoviridae*.

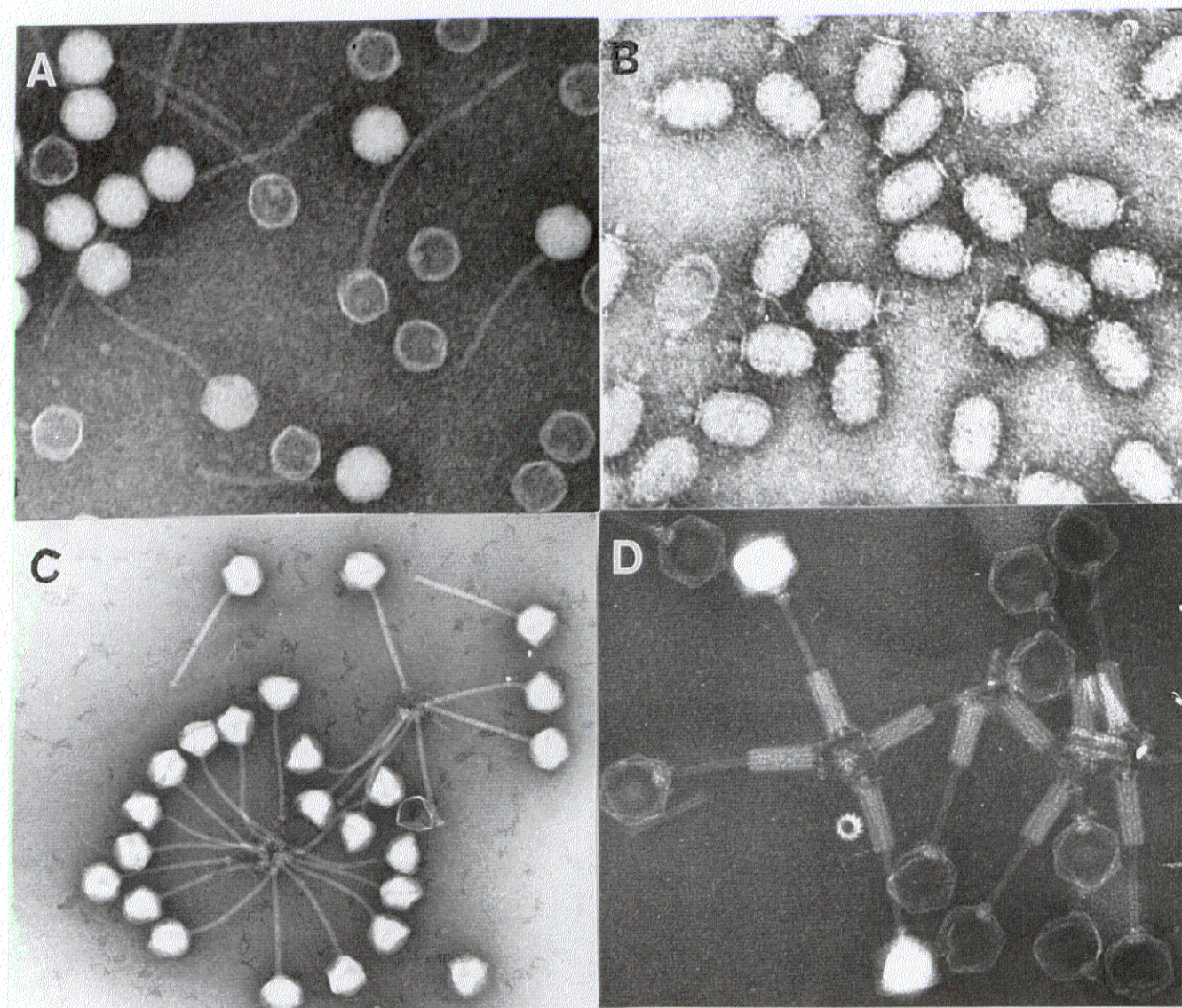


Figura 1. Fotografías de los fagos de neumococo tomadas al microscopio electrónico. **A)** Dp-1; 125.000 aumentos. **B)** Cp-1; 175.000 aumentos. **C)** HB-3; 83.000 aumentos. **D)** EJ-1; 170.000 aumentos.

La especificidad de huésped de algunos de estos fagos no es absoluta ya que, en el caso de Cp-1, se ha demostrado que también se multiplica en *S. oralis*, una especie que, como neumococo, habita en el tracto respiratorio superior (Ronda y cols., 1989).

En la Tabla 1 se muestran una serie de características fisicoquímicas de los DNAs y de algunas de las proteínas de los fagos de neumococo. Entre las propiedades más destacadas de estos bacteriófagos, cabe señalar que Dp-1 es el primer fago de bacterias Gram-positivas en el que se describió la presencia de lípidos en sus estructuras periféricas (López y cols., 1977) y que codifica una enzima lítica que participa en la liberación de la descendencia fágica (Ronda y cols., 1977). Se han observado sólo pequeñas diferencias morfológicas entre los fagos Dp-1 y ω y los fagos atemperados HBs. Las cuatro familias de fagos, Dps, HBs, Cps y EJ-1 poseen DNAs lineales. Una característica peculiar que comparten sólo los miembros de la familia Cps y HBs es la presencia de una proteína unida covalentemente a los extremos 5' de sus DNAs, con la particularidad de que los fagos HBs, al ser atemperados, deben perder esa proteína durante su ciclo lisogénico para integrarse en el cromosoma del huésped (García y cols., 1983c; Romero y cols., 1990c).

2.2. El bacteriófago Cp-1 de *S. pneumoniae*

En el año 1981 se aislaron a partir de frotis de garganta tomadas de niños sanos en Alcalá de Henares (Madrid) una serie de fagos que infectaban a neumococo, denominados Cps. Todos ellos presentaban una morfología similar, aunque con algunas diferencias de tamaño. Cp-1, el representante más estudiado de esta familia de fagos, tiene una cabeza irregular hexagonal de aproximadamente 60×45 nm y una cola corta de casi 20 nm de longitud y 15 nm de grosor. La cabeza, con proyecciones, muestra una base aplastada y apéndices en el cuello (Ronda y cols., 1981). Esta morfología es muy similar a la que presenta el bacteriófago $\phi 29$, que infecta a *Bacillus subtilis*, otra bacteria Gram-positiva. Los fagos Cps contienen un DNA lineal de doble cadena con una proteína unida covalentemente a través de un residuo de treonina a sus extremos 5' (García y cols., 1986b). Esta proteína recibe el nombre de proteína terminal (TP) (García y cols., 1983c). En la clasificación de Ackermann y cols. (1995), Cp-1 está incluido en el grupo C2.

Tabla 1. Características de los fagos de neumococo, Dp-1, Cp-1, HB-3 y EJ-1

Características	Fago tipo			
	Dp-1	Cp-1	HB-3	EJ-1
<u>Familia</u>	<i>Siphoviridae</i>	<i>Podoviridae</i>	<i>Siphoviridae</i>	<i>Myoviridae</i>
Otros miembros	Dp-4, ω-1, ω-8	Cp-5, Cp-7, Cp-9	HB-746; HB-623	
<u>Morfología</u>				
Diámetro de cabeza (nm)	60	60 x 45	65	57
Longitud de la cola (nm)	155	20	156	130
Apéndices y fibras	No	Sí	No	Sí
Presencia de envuelta	Sí	No	No	No
Principales proteínas estructurales (kDa)	38 25	39 29,5	34 22	48 36
<u>Genoma fágico</u>				
Tipo de DNA	DS	DS	DS	DS
T_m (°C)	83,5	87	ND	ND
Tamaño (kb)	56	19	40	42
<u>Transfección</u>	Sí	Sí	No	No
<u>Proteína terminal unida al DNA (kDa)</u>	No	Sí (26,8)	Sí (23)	No
<u>Gen lítico (nt)</u>	ND	<i>cpl1</i> (1.017)	<i>hbl3</i> (954)	<i>ejl</i> (948)
<u>Enzima lítica (kDa)</u>	Amidasa (31)	Lisozima (39)	Amidasa (36)	Amidasa (36)
<u>Presencia de holina (kDa)</u>	ND	Sí (15,4)	Sí (12,9)	Sí (10,9)

DS : DNA bicatenario lineal.

ND : No determinado.

2.2.1. Propiedades biológicas

Las placas producidas por Cp-1 son más claras y grandes que las descritas para el resto de fagos de neumococo. También difiere del resto de estos fagos en su mayor estabilidad de conservación en diferentes condiciones experimentales. Esto, junto con su pequeño tamaño hacen de él una buena herramienta de trabajo para estudios genéticos. El bacteriófago Cp-1 se adsorbe muy lentamente a las células de neumococo (Ronda y cols., 1981) y su DNA se replica por un mecanismo denominado "protein-priming", en el que se usa la TP como donadora del grupo -OH para iniciar la replicación (García y cols., 1986b). La temperatura óptima de replicación es 30°C, observándose *in vivo* una inhibición casi completa a 37°C, mientras que *in vitro* esta inhibición es sólo parcial. En presencia de afidicolina, inhibidor de las DNA polimerasas α de células eucariotas, se inhibe la replicación de Cp-1, es decir, se bloquea la replicación continuada del DNA de Cp-1 sin afectar la formación del complejo de iniciación. Por otra parte, no se observa un bloqueo de la replicación de Cp-1 en presencia de HPUra, sustancia que si impide la replicación del DNA de la célula huésped (García y cols., 1986c).

La presencia de la TP es imprescindible para demostrar la actividad transfectante del DNA del fago Cp-1, que a su vez, se estimula por iones de calcio. Asimismo, se demostró que la desaparición de la TP unida covalentemente al DNA no afecta a las primeras etapas de incorporación del DNA (adsorción y entrada) (Ronda y cols., 1983), lo que sugiere que, al igual que en el caso de $\phi 29$, la TP de Cp-1 interviene en la replicación del DNA lineal de este virus.

2.2.2. Propiedades del DNA y de las proteínas del virión

Como ya se ha indicado, el ácido nucleico de Cp-1 es un DNA lineal de doble cadena con una longitud media de 6,3 μm , y un coeficiente de sedimentación de 19S, que se corresponde con un tamaño de alrededor de 12×10^6 Da. La temperatura de fusión (T_m) del DNA de Cp-1 es 87°C, que indica un contenido en G + C de 42%. La densidad de la doble banda del DNA en gradientes neutros de cloruro de cesio es 1,699 g/cm³ (similar a la del DNA del huésped). Esta densidad se correlaciona con un contenido en G + C de 40% que se corresponde con el dato obtenido de la temperatura de fusión del DNA. Por otra parte, la

densidad del DNA de este fago en gradientes de sulfato de cesio es $1,422 \text{ g/cm}^3$ indicando que no existen bases modificadas.

El DNA de Cp-1 no es digerido por un gran número de enzimas de restricción (Ronda y cols., 1981). Asimismo, la determinación de la secuencia de nucleótidos (nt) de los extremos de su genoma mostró una repetición terminal invertida (ITR) de 236 pb y los 116 nt siguientes presentan una similitud del 93% entre ambos extremos (Escarmís y cols., 1984). Estas ITRs están presentes también en otros fagos Cps: Cp-5 presenta una ITR de 343 pb y Cp-7 de 347 pb (Escarmís y cols., 1985). La similitud global entre las ITRs de estos tres fagos varía entre un 84% y un 92%, lo que se aproxima a la similitud total de los genomas de los fagos previamente determinada a través de análisis de restricción (López y cols., 1984). Los 39 primeros nt son idénticos en los tres fagos de neumococo, concentrándose las principales diferencias entre los nt 74 y 98. La conclusión que se saca de estos análisis es que los fagos Cp-1 y Cp-5 estaban más relacionados entre sí que con el fago Cp-7. Otros DNAs con TP covalentemente unida a sus extremos 5' también presentan ITRs de longitud variable, como veremos en el apartado 3.

La densidad del virión es $1,46 \text{ g/cm}^3$ y los análisis de los viriones en geles de poliacrilamida-SDS revelan la presencia de nueve proteínas estructurales, de las cuales una representa aproximadamente el 90% de las proteínas totales, con un M_r de 37.000. (Ronda y cols., 1981).

3. REPLICACIÓN DEL DNA

La replicación es el proceso a través del cual una molécula de DNA se obtienen dos copias idénticas a la original. Cada una de las cadenas sirve de molde para la nueva molécula a sintetizar y el sentido de crecimiento de la nueva cadena es siempre $5' \rightarrow 3'$. Este proceso se realiza con la participación de unas enzimas llamadas DNA polimerasas. Todas las DNA polimerasas conocidas necesitan un grupo $3'\text{-OH}$ libre para poder elongar la cadena, lo cual plantea el problema de cómo se inicia la replicación.

Generalmente, los grupos $3'\text{-OH}$ necesarios para iniciar la replicación son proporcionados por cadenas cortas de RNA sintetizadas por una proteína llamada "primasa" (Kornberg, 1980). Estas pequeñas moléculas de RNA son eliminadas tras la replicación y los huecos resultantes en la cadena de síntesis discontinua son rellenados por DNA

polimerasas, utilizando como iniciador el extremo 3'-OH del fragmento anterior de DNA. El uso de estos pequeños RNAs como cebadores de la iniciación de la replicación puede ser utilizado en el caso de moléculas circulares de DNA, o lineales que se convierten en circulares o en concatémeros en el momento de la replicación. Una alternativa a este mecanismo para la iniciación de la replicación, es el corte específico de una de las cadenas de la doble banda de DNA circular, liberándose un grupo 3'-OH que se usa para la elongación de la cadena naciente (Kornberg, 1980).

Para la replicación de los extremos de los genomas lineales, que permanecen como tales durante todo su ciclo replicativo, no es posible utilizar el mecanismo de iniciación con RNA, porque tras la eliminación de las moléculas iniciadoras, la DNA polimerasa no podría rellenar los huecos resultantes (Salas, 1988). Algunas moléculas de DNA lineal contienen una secuencia palindrómica en sus extremos 3', lo cual les permite la formación de una estructura en horquilla que les provee del 3'-OH para la elongación (Kornberg, 1980, 1982; Kornberg y Baker, 1992). Por otra parte, en los genomas lineales de muchos organismos eucariotas, las telomerasas se encargan de replicar los extremos, empleando como molde una molécula de RNA acoplada al enzima (revisado por Blackburn, 1992).

La caracterización de proteínas específicas unidas covalentemente a los extremos 5' de los DNA lineales de algunos virus y plásmidos (TP), permitió la demostración de la existencia de un nuevo mecanismo para la iniciación de la replicación, que se conoce con el nombre de "protein-priming" o iniciación con proteína, ya citado anteriormente (revisado por Salas, 1991).

3.1. Mecanismo de iniciación con proteína o "protein-priming"

Este mecanismo fue propuesto por Rekosh y cols. (1977) para solucionar el problema en la iniciación de la replicación del DNA de adenovirus. En este mecanismo, el grupo 3'-OH libre necesario para iniciar la replicación no lo aportan moléculas de DNA ni de RNA, sino el grupo -OH de un residuo de serina, tirosina o treonina de la TP. La replicación de estos DNA lineales con TP ocurre por un mecanismo simétrico: la replicación comienza en los dos extremos del DNA y, por tanto, las dos cadenas se sintetizan de modo continuo. Este mecanismo ha sido estudiado en detalle en el caso de adenovirus, del bacteriófago $\phi 29$ de *B. subtilis* y de PRD1, un fago que infecta a *Escherichia coli* y otras

bacterias Gram-negativas como *Salmonella typhimurium* (Salas, 1988, 1991; Stillman, 1983).

En la Figura 2 podemos ver un esquema del mecanismo descrito para el fago $\phi 29$, en el que se distinguen tres etapas:

Etapas de iniciación.- Unas proteínas específicas de iniciación interaccionan con el origen de replicación, produciendo la apertura de la doble hélice y la exposición de una región de cadena simple. Una molécula libre de TP forma un complejo con la DNA polimerasa, y este complejo interacciona con los orígenes de replicación mediante el reconocimiento de la TP parental unida al DNA y secuencias específicas de los extremos del DNA. A continuación, la DNA polimerasa en presencia del dNTP correspondiente, cataliza la unión covalente entre el dNMP y el grupo -OH de un residuo específico de serina, tirosina o treonina de la TP.

Etapas de elongación.- A continuación, la misma DNA polimerasa replica procesivamente una cadena del DNA mientras produce el desplazamiento de la otra, que es protegida por una proteína de unión al DNA de cadena sencilla (SSB). Se pueden distinguir dos tipos de intermedios replicativos, que corresponden a la estructura de la molécula en replicación antes y después del encuentro entre las dos horquillas (revisado por Salas, 1991). Como se ha señalado anteriormente, la replicación es continua, y no son necesarias otro tipo de proteínas para completar la síntesis del DNA en toda su longitud a excepción, en el caso de adenovirus, de una actividad topoisomerasa I.

Etapas de terminación.- La DNA polimerasa se disocia de la molécula recién sintetizada y se une a otra TP libre para iniciar una nueva ronda de replicación.

3.1.1. Genomas con TP

Se han descrito varios ejemplos de DNAs lineales unidos covalentemente a una TP, son los casos de adenovirus y $\phi 29$ (a través de un residuo de serina de la proteína) (Hermoso y cols. 1985; Hermoso y Salas, 1985; Smart y Stillman, 1982) y PRD1 que se une por una tirosina (Bamford y Mindich., 1984). Otros DNAs lineales con TP son los plásmidos pSLA1 y pSLA2 de *Streptomyces rochei* (Hirochika y cols., 1982, 1984), el plásmido pSCL de *Streptomyces clavuligerus* (Keen y cols., 1988), los plásmidos "killer" de levaduras pGKL1 y pGKL2 (Kikuchi y cols., 1984) y los DNAs mitocondriales S1 y S2 de

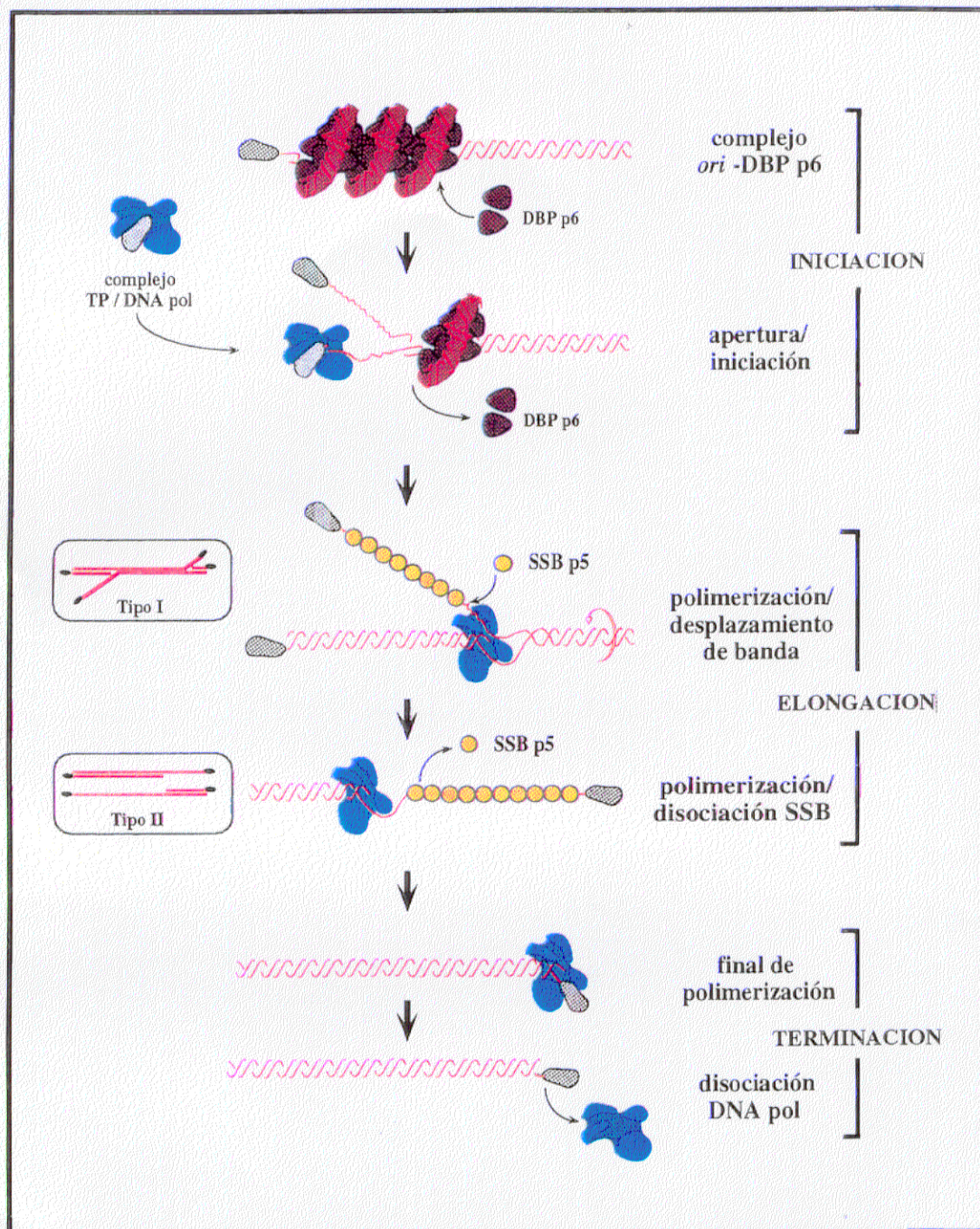


Figura 2. Mecanismo de replicación del DNA del bacteriófago $\phi 29$. Este mecanismo se comenta con detalle en el texto. El esquema está tomado de la Memoria de Tesis Doctoral de J. Méndez (1995).

maíz (Kemble y Thompson, 1982; Paillard y cols., 1985), entre otros. Además, también se han caracterizado TPs unidas a extremos 5' de RNA de virus animales (Daubert y Bruening, 1984). Recientemente se ha comprobado que cromosomas bacterianos como los de *Borrelia* o *Streptomyces* también son lineales y poseen TP unida covalentemente a sus extremos (Hinnebusch y Tilly, 1993; Lin y cols., 1993). A pesar de que el mecanismo de "protein-priming" se postula como un mecanismo universal de replicación para los DNAs con TP, en 1994 se demostró que la replicación del DNA del plásmido pSLA2 tiene lugar a partir de un origen de replicación interno y el mecanismo que utiliza la TP como donadora del grupo 3'-OH queda reducido a la replicación de los extremos, y no a la totalidad del genoma (Chang y Cohen, 1994). Este mismo procedimiento se postula para la replicación de los cromosomas lineales con TP de bacterias del género *Streptomyces*.

Los DNAs con TP que inician la replicación utilizando como cebador el grupo -OH de un residuo de una proteína, presentan ITRs de diferente longitud, en el caso de $\phi 29$ de 6 pb (Escarmís y Salas, 1981) y en el caso de PRD1 y otros fagos relacionados con él de 109 y 110 pb. La presencia de la secuencia CCCCT(A)CCC en las ITRs de los fagos Cps y PRD1, pero sólo en el lado izquierdo del genoma de $\phi 29$, extremo por donde se empaqueta en primer lugar la partícula fágica, ha llevado a sugerir un papel fundamental de dicha secuencia en el proceso de empaquetamiento (Salas, 1988). Los diferentes tipos de adenovirus muestran unas ITRs de alrededor de 100 pb (Tolun y cols., 1979). En el caso de $\phi 29$, de adenovirus, y de PRD1, se ha demostrado que las ITRs contienen el origen de replicación, permitiendo que la síntesis del DNA empiece en los dos extremos de la molécula.

3.1.2. "Sliding-back": un nuevo mecanismo para la transición entre la iniciación y la elongación en los genomas con proteína terminal

La formación del complejo TP-dNMP en el paso de iniciación de la replicación es dirigida por complementaridad de bases con un nucleótido del DNA molde. Teniendo en cuenta que el mecanismo de la iniciación con proteína es un recurso evolutivo de estos genomas lineales para solucionar el problema de la replicación de los extremos del DNA, inicialmente se pensó que la iniciación de la replicación era dirigida por el primer nucleótido del extremo 3'. Después de un estudio mediante análisis mutacional de los tres primeros

nucleótidos del extremo 3' del DNA del fago $\phi 29$ se demostró que el nucleótido en segunda posición desde el extremo 3' es el que dirige la reacción de iniciación. Posteriormente mediante un análisis de replicación truncada se demostró que, a pesar de que la iniciación no ocurre en el extremo del DNA, la replicación es completa, es decir, la información presente en el nucleótido terminal se recupera. Además, las moléculas de DNA han de tener una reiteración de al menos dos nucleótidos en el extremo para ser replicados de forma eficiente. De esta manera, el mecanismo de transición entre la iniciación y la elongación permite la recuperación del nucleótido del extremo. Dicho mecanismo, descrito y caracterizado por primera vez en $\phi 29$, se ha denominado "sliding-back" (Méndez y cols., 1992).

Este mecanismo se basa en que la DNA polimerasa cataliza la adición de dAMP a la TP, dirigida por la segunda timina del molde. A continuación, el complejo TP-dAMP realiza una translocación hacia atrás ("sliding-back") de manera que el nucleótido incorporado a la TP se coloca frente a la primera timina. Entonces, la DNA polimerasa vuelve a utilizar el segundo nucleótido como molde para extender el complejo TP-dAMP, dando lugar a TP-(dAMP)₂. A partir de este momento, la elongación ocurre normalmente.

Las secuencias terminales de otros DNAs que utilizan proteínas como iniciadores de la replicación (bacteriófagos, virus de eucariotas y plásmidos lineales) presentan algún tipo de reiteración de secuencia en el extremo. La existencia de estas repeticiones terminales sugirió que este mecanismo podría darse también en otros sistemas que utilizan la proteína como donadora del grupo -OH para iniciar la replicación. Esta hipótesis se demostró posteriormente en el fago PRD1 (Caldentey y cols., 1993) y en los adenovirus (King y Van der Vliet., 1994). La replicación de PRD1 se inicia frente al cuarto nucleótido del DNA molde, lo que sugiere que el extremo se regenera mediante un mecanismo de "sliding-back" (Caldentey y cols., 1993). En el caso de adenovirus, el inicio de la replicación también es frente al cuarto nucleótido del molde, pero en este sistema se produce una pequeña variación respecto a los anteriores. Tras la iniciación en la cuarta posición, la DNA polimerasa catalizaría la formación de la molécula TP-CAT, que daría un "salto hacia atrás" para aparear con el primer trinucleótido del molde 3'-GTA, antes de que comience la elongación normal (King y Van der Vliet, 1994). Esta variación respecto del "sliding-back" se ha denominado "jumping-back".

4. MECANISMOS DE LISIS: LIBERACIÓN DE LA PROGENIE FÁGICA

La liberación de la progenie fágica de las células infectadas es la etapa final del proceso de infección. Todos los bacteriófagos, excepto los filamentosos, requieren la lisis del huésped para su liberación. A pesar de que en los libros de texto se mantiene como concepto clásico que la lisis producida por los fagos es el resultado inevitable de la acumulación de la enzima lítica fágica durante la síntesis de proteínas tardías resulta sorprendente, sin embargo, que haya muchos fagos, como ϕ X174 y MS2 que no codifican ninguna enzima lítica y que, por otra parte, en otros fagos como lambda, T4 y posiblemente T7, la acumulación de la actividad lítica es independiente del momento en que se produce la lisis. Estas ideas contrapuestas pueden ser debidas a que los estudios sobre la lisis fágica han progresado lentamente y no se han podido elucidar con detalle las distintas etapas que tienen lugar durante la lisis. Esto se debe a que el análisis de los genes líticos es una tarea complicada debido fundamentalmente a que estos genes están diseñados para matar a la bacteria, en un momento determinado, y en condiciones fisiológicas se expresan a niveles moderados, además de participar en la continua remodelación de la pared bacteriana; por tanto, su clonación en plásmidos multicopia requiere una estricta regulación, lo que conlleva ciertas dificultades. Además, las demás proteínas implicadas en la lisis, aparte de las enzimas líticas, son difíciles de purificar, debido en parte a su pequeño tamaño, su ubicación en la membrana y a su carencia de actividad enzimática, lo cual dificulta su manipulación. Además de todo esto, no se pueden hiperexpresar ya que son letales para las células. Otro de los aspectos críticos del análisis del fenómeno lítico es el estudio de su regulación, ya que no es suficiente identificar los genes implicados en la lisis, sino que se debe intentar explicar cómo ocurre la lisis en el momento preciso. La lisis puede ser ineficaz tanto si no se produce, con lo que no hay liberación de la descendencia fágica, como si se produce demasiado pronto antes de que se formen los viriones infectivos. En cuanto a la regulación del momento correcto en que se produce la lisis hay que destacar que existe un control a nivel de traducción y otro a nivel de la función de la proteína.

5301691857

4.1. Papel de las holinas y de las enzimas líticas en el proceso de lisis

4.1.1. Holinas

Se han descrito al menos dos mecanismos diferentes a través de los cuales los fagos provocan la lisis de la célula huésped (Young, 1992). Algunos fagos pequeños han desarrollado un mecanismo de lisis mediante un único gen autolítico, cuyo producto no posee actividad lítica. En el caso del fago ϕ X174 que infecta a *E. coli* se ha demostrado la existencia de un gen responsable de la lisis celular y de la liberación de la descendencia fágica que parece actuar como activador del sistema autolítico de la célula huésped (Bläsi y cols., 1984). Muchos otros fagos, sin embargo, codifican una actividad lítica de pared, producto de un gen tardío, que se sintetiza como una proteína soluble en el citoplasma. Esta enzima alcanza su sustrato en el periplasma mediante la acción de una segunda proteína codificada por el fago, la holina, nombre derivado del término inglés “hole”. Estas proteínas han recibido el nombre de holinas porque su función es la de hacer “agujeros” en la membrana citoplasmática (Young, 1992), y para poder diferenciarlas de la familia de otras proteínas de membrana llamadas porinas. Además, las lesiones producidas por las holinas en la membrana no son específicas, es decir, que permiten el paso de cualquier enzima lítica sintetizada en el citoplasma. Este sistema dual de lisis ha sido caracterizado detalladamente en el bacteriófago lambda y otros fagos relacionados.

El operón que controla la lisis del fago lambda comprende tres genes tardíos: *S*, que codifica la holina, *R*, que codifica una transglicosilasa específica del peptidoglicano de *E. coli* y *R₂*, que codifica una proteína de función desconocida, que no es esencial para la lisis excepto en condiciones de altas concentraciones de cationes divalentes. *R₂* es una proteína de membrana cuya secuencia N-terminal es muy similar a las típicas señales secretoras. El producto del gen *S* se oligomeriza en la membrana interna para formar “agujeros” inespecíficos que permiten el paso de la transglicosilasa al periplasma (Zagotta y Wilson, 1990). La funcionalidad de esta proteína se ha demostrado incluso en *Saccharomyces cerevisiae* donde se ha comprobado que forma oligómeros similares a los observados en la membrana de *E. coli*. En este sentido, la proteína *S* no requiere una diana en la membrana del huésped para el anclaje del oligómero, por lo que se puede considerar como una citolisina interna capaz de actuar en células eucariotas como un factor letal (Garrett y cols.,

1990). La expresión de *S*, incluso en ausencia de otros genes líticos, es letal para la célula huésped puesto que produce un aumento de permeabilidad en la membrana plasmática, colapsa el potencial de membrana, inhibe la respiración y el transporte activo es defectivo (Garrett y Young, 1982; Young, 1992). La ausencia de la proteína *S* hace que la transglicosilasa se acumule intracelularmente, la lisis celular no tiene lugar y la respiración y metabolismo continúan una vez pasado el tiempo normal en el que se produciría la lisis. Por el contrario, en presencia de *S* y ausencia de la transglicosilasa no se produce lisis pero la célula está fisiológicamente muerta. El momento en que se produce la lisis depende totalmente de la función de la proteína *S*. No obstante, el gen *S* codifica dos proteínas, una que empieza a traducirse a partir de la metionina 1 (S107) y otra a partir de la metionina en posición 3 (S105), y ambos codones de iniciación ATG, están precedidos por un sitio de unión al ribosoma. Esta característica se ha llamado motivo de inicio doble de traducción.

El gen *S* se transcribe constitutivamente a partir de los 30 min después de la infección; por lo tanto no hay evidencias de regulación a nivel de la transcripción durante el ciclo vegetativo. Se ha demostrado en varios sistemas, lambda y fagos lambdoides (Bläsi y cols., 1989; Bonovich y Young, 1991) y P22 (Nam y cols., 1989) entre otros, que los dos codones de metionina son utilizados *in vivo*, y que las dos proteínas de diferente tamaño que se forman tienen funciones opuestas. La de mayor tamaño (S107 en lambda) actúa como regulador negativo de la otra (S105 en lambda), que es la que propiamente ejerce la función letal para el huésped (Bläsi y cols., 1989). La única diferencia entre las dos especies es la presencia de, al menos, una carga positiva entre los dos residuos de metionina. La relación *in vivo* entre las dos proteínas se mantiene en la proporción 2:1 en favor de S105 (Bläsi y cols., 1989). Por lo tanto, puesto que ambas especies tienen funciones opuestas, y la relación de las dos formas se mantiene constante a lo largo del ciclo vegetativo, la regulación temporal de la formación del "agujero" no parece ser ni a través de un diferente nivel de síntesis de mRNA a partir de cada uno de los codones durante el ciclo vegetativo, ni a nivel traduccional. Por consiguiente, debe tratarse de una regulación a nivel postraduccional.

El mecanismo de inicio doble de traducción también se ha demostrado en fagos de bacterias Gram-positivas, como en $\phi 29$. En este caso, también la molécula más larga actúa como inhibidor de la verdadera holina que tiene dos aminoácidos menos (Steiner y cols., 1993; Tedin y cols., 1995).

Se ha propuesto un modelo de regulación basado en cambios conformacionales de las proteínas provocados por la destrucción del potencial de membrana. Primero, las dos formas se insertarían y se acumularían en la membrana a través de sus dominios hidrofóbicos transmembrana formando mezclas oligoméricas. Ambas proteínas se ensamblarían en la membrana aleatoriamente sin formar un “agujero” definido, en un proceso independiente de las funciones celulares *sec* (funciones de transporte) (Rietsch y Bläsi, 1993). En un segundo paso la disipación del potencial de membrana permitiría el cambio conformacional necesario para que se produzca el ensamblaje correcto para formar el “agujero” funcional (Bläsi y cols., 1990). La destrucción del potencial de membrana se puede producir por un incremento en la salida de protones producido por la acumulación de S105 y S107 (Steiner y Bläsi, 1993). Como se ha citado anteriormente, los productos del gen S tienen funciones opuestas, pero además, S107, tiene una capacidad doble. S107 funciona como un inhibidor mientras la célula infectada mantenga el potencial de membrana, y contribuye activamente a la formación del poro cuando se colapsa éste (Bläsi y cols., 1990; Steiner y Bläsi, 1993).

Las enzimas líticas de los fagos que se han caracterizado hasta hoy, están sintetizadas sin una secuencia señal que les permita llegar al peptidoglicano, y se van acumulando en el citoplasma durante el ciclo vegetativo. La única excepción conocida hasta el momento es la del profago defectivo PBSX que presenta péptido señal aunque, aparentemente, se exporta por un mecanismo dependiente de holina sin que se hidrolice dicho péptido (Longchamp y cols., 1994). Por tanto, las holinas, como se ha explicado anteriormente, son las encargadas de permitir la liberación de las enzimas líticas acumuladas en el citoplasma a la pared celular al final del ciclo vegetativo. A continuación el peptidoglicano se degrada y se produce la lisis celular.

Las holinas, a pesar de que a nivel de estructura primaria no presentan entre sí ninguna similitud, comparten una serie de características comunes. Son proteínas de bajo peso molecular con varios dominios transmembrana unidos por giro β , un C-terminal muy cargado e hidrofílico y un N-terminal también hidrofílico. En la mayoría de los casos conocidos, los genes que codifican las holinas preceden a los que especifican las enzimas líticas (Young, 1992; Young y Bläsi, 1995). Teniendo en cuenta esta serie de características, se ha podido elaborar una lista de genes que codifican holinas o posibles holinas (aquellas que todavía no se ha demostrado que funcionen como tales pero que

presentan las características estructurales comunes a todas ellas), en una gran variedad de fagos tanto de bacterias Gram-negativas como de Gram-positivas (Tabla 2). En la mayoría de los casos se cuenta además con alguna evidencia adicional, genética o fisiológica sobre la implicación del gen de la holina en la lisis del huésped. Young y Bläsi (1995) consideraron que todas estas proteínas son de hecho holinas funcionales y distinguieron tres tipos de holinas en fagos de bacterias Gram-negativas: tipo I (S de λ , gp13 de P22, Y de P2, LydA de P1), que tienen 87 residuos como mínimo, y contienen tres o más dominios que pueden ser considerados como dominios transmembrana; tipo II (S de 21, S de PA-2, S de ϕ 80, gp17.5 de T7, Lys de T3, ORF78 de Hp1), que tienen como máximo 78 aminoácidos y sólo presentan dos presuntos dominios transmembrana, y un tercer tipo formado por la proteína t de T4 (Young y Bläsi, 1995), que no comparte ninguna de las características descritas de la secuencia con cualquier otra holina aunque sí se ha demostrado que funciona como tal. Un fago T4 mutante para el gen *t*, produce la acumulación de la enzima lítica en el citoplasma sin producir la lisis celular (Josslin, 1970, 1971). La proteína t es mucho más larga que las otras holinas y sólo parece que presenta un dominio transmembrana. Por otra parte, recientemente se ha aislado y caracterizado en el cromosoma de *B. licheniformis* un gen que codifica una proteína funcional y estructuralmente similar a una holina y cuya secuencia presenta cierto grado de homología con la presunta holina propuesta para el profago PBSX (Kyogoku y Sekiguchi, 1996).

La topología de estas holinas en la membrana se puede predecir con más seguridad en las de tipo II. Los extremos N- y C-terminales cargados estarían dispuestos en el citoplasma y los dos dominios transmembrana embebidos en la bicapa lipídica como hélices antiparalelas. El dominio de unión entre los dos dominios hélice α , que forma una estructura de giro β , está cargado positivamente o tiene carga neta neutra (Von Heijne, 1992). La topología de las holinas de tipo I está menos establecida, especialmente la localización del extremo N- terminal; el extremo C-terminal se encuentra en el citosol.

Tabla 2. Holinas de fagos de bacterias Gram-negativas y positivas

Bacteriófago	M_r	Función	$Mx(x)M^a$	Tipo	Referencia
<u>Gram-negativas</u>					
<i>E. coli</i>					
λ	8.500	Sí	Sí	I	Altman y cols., 1983
21	7.800	Sí	Sí	II	Bonovich y Young, 1991
PA-2	7.800	Sí	Sí	II	Bonovich y Young, 1991
$\phi 80$	8.200	No	Sí	II	Young y Bläsi, 1995
P1	11.700	Sí	No	I	Schmidt y cols., 1996
P2	9.900	No	No	I	Ziermann y cols., 1994
HK22 ^b	--	--	Sí	--	Young y Bläsi, 1995
HK97 ^b	--	--	Sí	--	Young y Bläsi, 1995
T3	7.400	Sí	No	II	Miyazaki y col., 1978
T7	7.400	Sí	No	II	Yamada y cols., 1986
T4 ^c	25.000	Sí	-	-	Josslin, 1970, 1971
<i>E. coli</i> y <i>S. typhimurium</i>					
PRD1	9.800	No	No	I	Young y Bläsi, 1995
<i>S. typhimurium</i>					
P22	11.700	Sí	Sí	I	Rennell y Poteete, 1985
<i>Pseudomonas phaseolicola</i>					
$\phi 6$	6.000	Sí	?	?	Mindich y Lehman, 1979
<i>Haemophilus influenza</i>					
Hpl	8.700	No	No	II	Young y Bläsi, 1995
<u>Gram-positivas</u>					
<i>B. subtilis</i>					
$\phi 29$	14.800	Sí	Sí	I	Steiner y cols., 1993
					Tedin y cols., 1995
PZA	14.800	No	Sí	I	Paces y cols., 1986
$\phi 105$	14.000	No	Sí	I	Leung y Errington, 1995
PBSX ^d	10.000	No	No	I	Longchamp y cols., 1994
	9.700	No	No	I	
<i>S. pneumoniae</i>					
Ej-1	9.400	Sí	No	I	Díaz y cols., 1996
Cp-1	15.400	No	No	I	--
Hb-3	12.100	No	No	I	--
<i>Lactococcus lactis</i>					
$\phi LC3$	9.700	Sí	No	I	Birkeland, 1994
$\phi US3$	7.700	No	No	II	Platteuw y De Vos, 1992
Tuc2009	9.700	No	No	I	Arendt y cols., 1994
<i>Staphylococcus aureus</i>					
$\phi 11$	16.300	No	No	I	Young y Bläsi, 1995
$\phi vML3$	8.000	No	Sí	II	Shearman y cols., 1994
bIL67	10.800	No	No	I	Schouler y cols., 1994
<i>Lactobacillus delbrueckii</i>					
mv4	13.600	No	No	I	Dupont y cols., 1993
<i>Lactobacillus gasseri</i>					
ϕadh	12.900	Sí	No	I	Henrich y cols., 1995

Tabla 2. (continuación)

Bacteriófago	M_r	Función	$Mx(x)M^a$	Tipo	Referencia
<i>Lactobacillus casei</i> 392-A2	15.200	No	Si	I	Herrero, 1996
<i>Listeria monocytogenes</i> A518	10.100	No	Si	I	Loessner y cols., 1995
A500	10.300	No	Si	I	Loessner y cols., 1995
profagos crípticos	10.100	Sí	Si	I	Zink y cols., 1995

^a Indica la existencia del doble comienzo de la traducción implicado en la regulación de la función de la holina.

^b Datos no publicados.

^c Esta proteína no tiene las características estructurales de las holinas pero funciona como tal.

^d La acción combinada de ambas proteínas puede funcionar como una holina.

4.1.2. Enzimas líticas

Las enzimas líticas de pared hidrolizan uniones específicas de los peptidoglicanos, polímeros que forman parte de la pared bacteriana, solubilizando las paredes bacterianas y produciendo eventualmente la lisis y la muerte de la bacteria. Existen cuatro clases de enzimas líticas atendiendo a la especificidad del enlace hidrolizado (revisado por Rogers y cols., 1980):

i) Glicosidasas, que hidrolizan la cadena glicánica del peptidoglicano. Entre ellas se encuentran las *endo N*-acetilglucosaminidasas (glucosaminidasas), que liberan fragmentos con residuos de *N*-acetilglucosamina en el extremo reductor, y las *endo N*-acetilmuramididasas (muramididasas o lisozimas), que liberan fragmentos con *N*-acetilmurámico en el extremo reductor. ahora, incluyendo sus características bioquímicas más relevantes.

ii) *N*- acetilmuramil-*L*-alanil amididasas, que hidrolizan el enlace amida entre la cadena glicánica y el péptido.

iii) Endopeptidasas, que rompen la cadena peptídica del peptidoglicano, y de las cuales se han descrito varios ejemplos según el enlace hidrolizado dentro del péptido.

iv) Transglicosilasas, que liberan grupos no reductores de *N*-acetilmurámico.

Como se ha dicho anteriormente, con mucha frecuencia la hidrólisis de la pared bacteriana se lleva a cabo por enzimas líticas codificadas por el propio fago. Se han descrito

muchas enzimas líticas asociadas a fagos, pero sólo algunas han sido estudiadas en detalle. En la Tabla 3 se detalla la relación de algunas descritas hasta ahora, incluyendo sus características más relevantes.

Tabla 3. Enzimas líticas de fagos de bacterias Gram negativas y positivas

Bacteriófago	Función	M_r	Referencia
<u>Gram negativas</u>			
<i>E. coli</i>			
T4	lisozima	18.700	Tsugita e Inouye., 1968
T2	lisozima	18.600	Inouye y Tsugita, 1968
T3	amidasa	17.000	DeMartini y cols., 1975
T5	amidasa	17.000	DeMartini y cols., 1975
T7	amidasa	17.000	Inouye y cols., 1973
λ	transglicosilasa	17.500	Bienkowska-Szewczyk y Taylor, 1980
21	lisozima ^a	17.800	Bonovich y Young, 1991
PA-2	lisozima ^a	17.700	Bonovich y Young, 1991
P1	lisozima ^a	20.300	Schmidt y cols., 1996
P2	?	18.600	Ziermann y cols., 1994
<i>E. coli</i> y <i>S. typhimurium</i>			
PRD1	lisozima	17.300	Pakula y cols., 1989; Caldentey y cols., 1994
<i>S. typhimurium</i>			
P22	lisozima	16.100	Botstein y cols, 1972
<i>Pseudomonas phaseolicola</i>			
$\Phi 6$	murein hidrolasa	24.000	Mindich y Lehman, 1979
<i>Haemophilus influenzae</i>			
Hp1	lisozima ^a	?	Young, 1992
<u>Gram positivas</u>			
<i>B. subtilis</i>			
SF6	lisozima	34.700	Verma, 1986
$\phi 29$	lisozima	27.000	Saedi y cols., 1987
PZA	lisozima	27.000	Paces y cols., 1986
$\phi 105$	transglicosilasa	13.200	Leung y Errington, 1995
PBSX	amidasa	32.000	Longchamp y cols., 1994
	?	34.000	
PBSZ	?	33.000	Longchamp y cols., 1994
	?	34.000	
PBSY	?	32.000	Longchamp y cols., 1994
	?	33.000	
<i>Lactococcus lactis</i>			
P001	lisozima ^a	22.300	Geis, 1992; Sable y Lortal, 1995
C2	lisozima ^a	22.000	Ward y cols., 1993; Sable y Lortal, 1995
C10	?	?	Oram y Reiter, 1965; Sable y Lortal, 1995
C2(W)	lisozima	46.000	Mullan y Crawford, 1985; Sable y Lortal, 1995

Tabla 3. (continuación)

Bacteriófago	Función	M_r	Referencia
<i>Lactococcus lactis</i>			
φvML3	lisozima	20.000	Shearman y cols., 1994; Sable y Lortal, 1995
bIL67	lisozima ^a	25.700	Schouler y cols., 1994
φLC3	lisozima	46.300	Birkeland, 1994
Tuc2009	lisozima ^a	46.300	Arendt y cols., 1994
φUS3	amidasa ^a	29.000	Platteuw y De Vos, 1992
<i>Lactobacillus delbruckei</i>			
mv1	lisozima ^a	21.100	Boizet y cols., 1990
mv4	lisozima ^a	32.700	Dupont y cols., 1993
<i>Lactobacillus lactis</i>			
LL-H	?	?	Trautwetter y cols., 1986
<i>Lactobacillus casei</i>			
PL-1	lisozima	37.000	Hayashida y cols., 1987; Sable y Lortal, 1995
392-2A	amidasa ^a	33.400	Herrero, 1996
<i>Lactobacillus gasserie</i>			
φadh	lisozima ^a	34.700	Henrich y cols., 1995
<i>Listeria monocytogenes</i>			
A118	endopeptidasa	30.800	Loessner y cols., 1995
A500	endopeptidasa	33.400	Loessner y cols., 1995
A511	amidasa	36.500	Loessner y cols., 1995
φLM4		33.000	Payne y Gasson, 1990
profagos crípticos	?	33.000	Zink y cols., 1995
<i>S. pneumoniae</i>			
Dp-1	amidasa	31.000	García y cols., 1983a
Cp-1	lisozima	39.100	García y cols., 1987
Cp-7	lisozima	38.500	García y cols., 1990
Cp-9	lisozima	39.100	García y cols., 1990
HB-3	amidasa	36.000	Romero y cols., 1990a, b
EJ-1	amidasa	37.500	Díaz y cols., 1992a

^a Indica las proteínas que tienen actividad litica y por su similitud con otras secuencias de actividad demostrada se deduce el tipo de actividad que presenta.

Por otra parte, se ha comprobado que a veces las enzimas liticas intervienen en las etapas iniciales de la infección como es el caso del fago φ6, un virus RNA con envuelta lipídica que infecta a *Pseudomonas phaseolicola*, en el que se ha identificado una enzima litica que es una proteína estructural de la partícula fágica que facilita el paso del RNA fágico a través de las paredes bacterianas y que, posteriormente, junto con otra proteína del fago, participa en la lisis celular (Mindich y Lehman, 1979). También se han encontrado

funciones adicionales a la lítica en otras lisinas: las enzimas líticas de los bacteriófagos PZA y $\phi 29$ están implicadas en el empaquetamiento del DNA durante la morfogénesis de la partícula fágica (Camacho y cols., 1977; Jimenez y cols., 1977) y dada la alta similitud de secuencia de estas proteínas con la enzima lítica del fago $\phi vML3$, Shearman y cols. (1989) han postulado que esta proteína también presenta esta doble función (lítica y de empaquetamiento). A este respecto, se ha descrito otra proteína con funciones de maduración y lisis en el bacteriófago Q β (Karnik y Billeter, 1983; Winter y Gold, 1983).

Otro mecanismo de lisis, independiente de las proteínas fágicas, holina y enzima lítica, es el que presentan otros bacteriófagos como $\phi X174$ y MS2, en los que la liberación de la progenie fágica al final del ciclo lítico depende, como se ha señalado anteriormente, de las autolisinas de *E. coli*. La proteína E, codificada por el fago $\phi X174$, que causa la lisis de *E. coli* depende de varias proteínas, todas las cuales se encuentran asociadas a la pared celular. Otros parámetros como la composición lipídica, la fuerza protón-motriz de la membrana y la actividad del sistema autolítico influyen en la lisis celular mediada por el fago. Previamente a la lisis se producen alteraciones enzimáticas y estructurales tales como la inducción de las actividades fosfolipasa celulares, cambio en los patrones de proteína en las membrana externa e interna y una limitada degradación del peptidoglicano. Todos estos efectos son consecuencia directa o indirecta de la integración de la proteína E en la membrana interna y externa (Witte y Lubitz, 1989). Un importante aspecto de esta proteína codificada por el fago $\phi X174$ es que no posee actividad enzimática propia. A diferencia de lo que ocurre con las células lisadas por el fago lambda, las que han sido lisadas por $\phi X174$ aparecen vacías, pero mantienen su integridad estructural (Witte y Lubitz, 1989). La lisis se lleva a cabo por la formación de un "túnel" transmembrana realizado por la proteína E, que se oligomeriza para poder producir la lisis (Witte y cols., 1990). En ausencia de proteína E, la progenie fágica se acumula en el interior de la célula. Por el contrario, sólo su expresión es suficiente para producir la lisis de la célula. La proteína E es hidrofóbica y está dividida en cuatro dominios, cada uno de los cuales termina en un giro β . Aparte de las enzimas líticas del huésped, hay otros genes en la bacteria que están implicados en la lisis, aunque no son esenciales como los genes *slyD* (genes de sensibilidad a la lisis) (Young, 1992; Roof y Young, 1995). Schön y cols. (1995) han propuesto un modelo de integración de la proteína E en la membrana con el resultado final de la formación de un "túnel" mediante la fusión de la membrana interna y externa.

La lisis fágica en bacterias Gram-positivas ha sido menos estudiada. Como la autólisis en muchas bacterias Gram-positivas se induce espontáneamente bajo ciertas condiciones fisiológicas, se podría esperar que los fagos lisaran a sus huéspedes por inducción de las autolisinas. Sin embargo, un sistema similar al de lambda se ha detectado en $\phi 29$ (Steiner y cols., 1993), en el fago atemperado $\phi LC3$ de *Lactococcus lactis* (Birkeland, 1994), el fago ϕadh de *Lactobacillus gasseri* (Henrich y cols., 1995) y el fago atemperado EJ-1 de *S. pneumoniae* (Díaz y cols., 1996). Los genes de la holina y de la enzima lítica de estos fagos están dispuestos de manera similar a la que presentan los fagos lambdoides, es decir, el gen de la holina precede inmediatamente al de la enzima lítica. Además, las principales características que distinguen a las holinas de los fagos de bacterias Gram-negativas (estructura secundaria, perfiles hidropáticos, etc) están presentes también en los que infectan a Gram-positivas. Basados en la localización del gen y en la estructura de la proteína que codifica, se han propuesto otras posibles holinas en fagos de bacterias Gram-positivas (Tabla 2). Esto sugiere que el sistema dual de lisis estudiado detalladamente en fagos de bacterias Gram-negativas puede ser el mecanismo más común de liberación de la progenie fágica también en bacterias Gram-positivas.

4.2. Enzimas líticas de *S. pneumoniae* y sus bacteriófagos

La presencia de enzimas autolíticas en bacterias es un hecho prácticamente general. En el caso particular de neumococo, las conocidas “tendencias suicidas” de este microorganismo, propiciadas por la acción de su principal autolisina al final de la fase estacionaria de multiplicación, puede interpretarse como una vía para proporcionar, en ambientes naturales, DNA donador que sería usado en los procesos de transformación genética que se desarrollan en este sistema, de forma espontánea.

En *S. pneumoniae* se han descrito hasta la fecha dos enzimas autolíticas, la amidasa LytA, que fue la primera que se descubrió y la mayoritaria en este sistema (Howard y Gooder, 1974) y una glucosaminidasa (Sánchez-Puelles y cols., 1986b; García y cols., 1989). En cuanto a los bacteriófagos de neumococo se han podido caracterizar como amidasas las enzimas líticas de los fagos Dp-1, EJ-1 y HB-3, y como muramidasa las de los fagos Cp-1, Cp-7 y Cp-9.

4.2.1. Autolisinas de neumococo

LytA es la enzima lítica mayoritaria presente en neumococo, se ha caracterizado como una *N*-acetilmuramil-*L*-alanil amidasa (Howard y Gooder, 1974), y se ha purificado a homogeneidad electroforética (Höltje y Tomasz, 1976). La actividad de esta enzima es totalmente dependiente de la presencia de residuos de colina en los ácidos teicoicos de la pared celular (Höltje y Tomasz, 1975a, b). Tanto la eliminación de los ácidos teicoicos de la pared como la utilización de paredes celulares en las que la colina ha sido sustituida por su análogo etanolamina hacen a la enzima incapaz de hidrolizar la pared, desencadenando una serie de cambios drásticos en la morfología de neumococo. Hay que señalar que neumococo fue el primer microorganismo en el que se describió la presencia de colina como componente de los ácidos teicoicos de la pared y que presenta un requerimiento nutricional absoluto por este aminoalcohol, de tal manera que su eliminación del medio de cultivo provoca la detención de la multiplicación celular (Rane y Subbarow, 1940; Tomasz, 1967). Por otra parte, se ha demostrado que LytA hidroliza la pared de *S. oralis*, una especie taxonómicamente muy próxima a neumococo (Ronda y cols., 1988), que también posee colina en la pared (Kilpper-Bälz y cols., 1985), aunque para este microorganismo la colina no parece ser un nutriente esencial.

El aislamiento del gen *lytA* facilitó la caracterización de una serie de mutantes entre los que cabe destacar la estirpe M31 que carece completamente de este gen. La estirpe M31 permitió el estudio del papel biológico que desempeña la principal autolisina de neumococo (Sánchez-Puelles y cols., 1986a, b). Este mutante crece a 37°C formando pequeñas cadenas, no se lisa al final de la fase estacionaria de crecimiento, ni en presencia de detergentes y muestra una respuesta tolerante frente a los antibióticos β -lactámicos. Todas estas características desaparecen cuando en esta cepa se clona y se expresa en un plásmido adecuado el gen *lytA* (Ronda y cols., 1987). La conclusión de estos experimentos fue que la función fisiológica de LytA era la separación de las células hijas al final de la división celular, aunque no indicaba que fuera indispensable para adquirir el estado de competencia para desarrollar transformación genética ni alteraba su multiplicación celular. Posteriormente, se ha sugerido también que esta enzima representa uno de los principales factores de virulencia de neumococo (Berry y cols., 1989).

La expresión del gen *lytA* en *E. coli* permitió demostrar que la autolisina se sintetiza como una forma de baja actividad específica (forma E), equivalente a la que se obtiene en células de neumococo cultivadas en un medio con etanolamina y que puede ser convertida en una forma de alta actividad específica (forma C) en presencia de paredes de neumococo que contengan colina en sus ácidos teicoicos o en presencia de colina. Este proceso recibió el nombre de *conversión*.

El aislamiento del mutante M31, permitió la detección de una nueva actividad lítica que fue caracterizada como una glucosaminidasa (Sánchez-Puelles y cols., 1986a, b). La estirpe M31 es, por tanto, una cepa adecuada para la purificación de la nueva actividad lítica, paso esencial para la asignación de un papel biológico a esta nueva enzima. Dicha enzima se purificó a homogeneidad electroforética asignándole un peso molecular de 64.000 kDa (García y cols., 1989). Al igual que para la amidasa *LytA*, la colina de los ácidos teicoicos de pared es fundamental para la actividad de esta enzima. La glucosaminidasa es capaz de degradar paredes con colina pero no con etanolamina y su actividad también se inhibe en presencia de colina y de los ácidos lipoteicoicos (LTA). Su temperatura óptima de actuación es 30°C, produciendo la lisis de la cepa M31 al final de la fase estacionaria de crecimiento cuando se incuba a dicha temperatura (Sánchez-Puelles y cols., 1986b).

Aunque la glucosaminidasa de *S. pneumoniae* no provoque autólisis de los cultivos a 37°C, no se puede descartar que a esta temperatura la enzima ejerza una actividad lítica limitada pero importante para la célula que pudiese explicar su posible contribución al proceso de separación de las células hijas al final del ciclo de división celular. Se ha sugerido que las autolisinas presentes en cantidades minoritarias intervienen en el remodelado de la pared celular durante el crecimiento (Kawamura y Shockman, 1983), así como en la consecución del estado de competencia para transformación genética (Lacks y Neuberger, 1975). La respuesta definitiva a estos interrogantes pasa por la obtención de mutantes defectivos en esta segunda autolisina.

4.2.2. Enzimas líticas de los bacteriófagos de *S. pneumoniae*

La primera enzima lítica fágica capaz de hidrolizar la pared celular de neumococo que fue aislada y purificada a homogeneidad electroforética fue la codificada por el fago Dp-1, y recibió el nombre de PAL (Phage-Associated Lysin) (García y cols., 1983a). Esta

enzima tiene una M_r de 31.000 y se ha caracterizado como una *N*-acetilmuramil-*L*-alanil amidasa (García y cols., 1984). Para llevar a cabo su actividad catalítica requiere también la presencia de residuos de colina en la pared y se inhibe en presencia de LTA. Sin embargo, a diferencia de *LytA*, su actividad se estimula por agentes reductores como el DTT o el β -mercaptoetanol (García y cols., 1983d), se inhibe parcialmente por cardiolipina y no da reacción inmunológica cruzada con la forma E de la amidasa *LytA*, aunque sí con la forma C (García y cols., 1984). Por último, se ha demostrado que esta enzima interviene en la liberación de la descendencia fágica en mutantes de neumococo que carecen de amidasa *LytA* por lo que ambas enzimas estarían implicadas en dicho proceso (Ronda y cols., 1977; García y cols., 1983b).

La segunda enzima lítica caracterizada fue la muramidasa o lisozima del fago Cp-1. Fue purificada a homogeneidad electroforética y tiene un M_r de 39.000. Esta enzima no necesita conversión para alcanzar su máxima actividad, pero si requiere la presencia de colina en los ácidos teicoicos de la pared. La colina libre y el LTA a altas concentraciones inhiben de manera no competitiva la actividad de la enzima (García y cols., 1987). Las similitudes bioquímicas y funcionales entre la amidasa del huésped y la lisozima codificada por el fago Cp-1, sirvieron de punto de referencia para proponer que los genes que codifican estas enzimas podrían ser similares en parte a nivel de nucleótidos. Esta hipótesis se analizó experimentalmente por medio de ensayos de hibridación (García y cols., 1988) que permitieron la clonación del gen *cpl1* del fago Cp-1 (García y cols., 1987). Para conseguir un mejor conocimiento de las características moleculares de los genes líticos presentes en otros fagos que infectan neumococo se llevó a cabo la clonación y secuenciación de los genes líticos *cpl9* del fago Cp-9 y *cpl7* del fago Cp-7. Para ello, se utilizaron como sonda en las hibridaciones los genes *lytA* y *cpl1*, respectivamente en condiciones poco restrictivas (García y cols., 1990). De la comparación de la estructura primaria de estos genes se pudo observar una mayor similitud entre los genes *cpl1* y *cpl9* que con el gen *cpl7*. Los genes *cpl9* y *cpl7* codifican proteínas de 342 y 349 aminoácidos con M_r de 38.461 y 39.107, respectivamente. *Cpl9* es una lisozima dependiente de la presencia de colina para su actividad y con dominio C-terminal homólogo al de la amidasa *LytA* y la lisozima *Cpl1*. Por el contrario, en el caso de la lisozima codificada por el fago Cp-7, se vió que era capaz de degradar, *in vivo* e *in vitro*, tanto paredes de neumococo con colina como paredes en las que la colina había sido reemplazada por su análogo estructural

etanolamina. El dominio C-terminal de Cpl7 era completamente diferente a los encontrados en las enzimas líticas anteriores. La similitud de los C-terminales de todas estas enzimas líticas con excepción de Cpl7 indica que este dominio es el responsable de la unión al sustrato.

Por otra parte, también se ha clonado y secuenciado el gen lítico *hbl3*, del fago atemperado HB-3 (Romero y cols., 1990a, b). La proteína que codifica es una amidasa dependiente de colina y, al igual que la LytA, requiere el proceso de conversión. La comparación de nucleótidos de ambos genes proporcionó uno de los ejemplos de mayor similitud entre genes de un fago y una bacteria (87%), similitud que desaparece en las regiones flanqueantes de dichos genes.

Más recientemente se ha clonado y secuenciado el gen lítico del fago EJ-1 (Díaz y cols., 1992a), otro fago atemperado aislado de una cepa clínica y atípica de neumococo. Este gen codifica una amidasa (Ejl) que depende para su actividad de la presencia de colina en el sustrato que degrada y requiere el proceso de conversión para conseguir su máxima actividad. En contraste con la amidasa LytA del huésped, en el caso de Ejl el proceso de conversión puede ser reversible por diálisis. La amidasa Ejl presenta una M_r de 36.535 y posee un dominio C-terminal homólogo al de la amidasa LytA. Presenta una similitud del 80% con la amidasa LytA en la región estructural, existiendo una divergencia completa en las zonas flanqueantes. Tanto Ejl como Hbl3 se inhiben, al igual que la amidasa del huésped, con las mismas concentraciones de LTA o colina.

Todos los dominios C-terminales de las enzimas líticas de neumococo y sus fagos, excepto en el caso de Cpl7, están constituidos por seis motivos de secuencias repetidas, de unos 20 aminoácidos de longitud cada uno, que presentan una secuencia consenso. Esta estructura primaria de los C-terminales en todas las enzimas líticas estudiadas y su dependencia de colina, ha llevado a sugerir que este aminoalcohol ha servido a lo largo del proceso evolutivo como un elemento de fuerte presión selectiva que ha contribuido a la conservación de las regiones C-terminales de todas estas enzimas líticas (García y cols. 1988). El dominio C-terminal de Cpl7 presenta 2,8 repeticiones exactamente iguales, de 48 aminoácidos cada una, y diferentes a las anteriores. Esto indica también una organización modular en su región C-terminal pero distinta a la del resto de las enzimas líticas del sistema de neumococo.

Por otra parte, el análisis global de las enzimas líticas de neumococo y sus fagos permitió deducir que las enzimas que presentan un mismo tipo de actividad, contenían dominios N-terminales semejantes. Así, el dominio N-terminal de la amidasa LytA de neumococo era homólogo al de las amidasas Hbl3 y Ejl, y el dominio N-terminal de la lisozima Cpl-1 al de las lisozimas Cpl7 y Cpl9 (García y cols., 1990). Esto apoya la hipótesis de que los dominios N-terminales de las enzimas líticas pueden haber constituido a lo largo de la evolución proteínas independientes. El hecho de que los dominios N-terminales presenten una débil actividad catalítica cuando se expresan de forma independiente, en comparación con las enzimas completas, sugiere que la adquisición del dominio C-terminal representa una ventaja evolutiva para aquellas enzimas que interactúan con sustratos poliméricos, como es el caso de la pared bacteriana, lo que permitiría una mejora en su eficiencia catalítica (López y cols., 1992, 1994).

Tomados en su conjunto los datos anteriores nos llevan a sugerir la existencia de una organización modular de las enzimas líticas del sistema de neumococo, donde el dominio C-terminal estaría implicado en funciones de reconocimiento y unión al sustrato mientras que el centro activo se localizaría en el dominio N-terminal. Esta hipótesis se ha demostrado posteriormente con la construcción de proteínas quiméricas entre las enzimas líticas de los distintos fagos estudiados y la amidasa del huésped mediante el intercambio de los dominios N- y C-terminales (Díaz y cols., 1990, 1991). La construcción de una enzima quimérica entre el dominio C-terminal de la lisozima del fago Cp-1 y el N-terminal de la enzima lítica de *Clostridium acetobutylicum* ATCC 824 (Croux y cols., 1993a), así como la más reciente construcción de una enzima híbrida construida entre el dominio N-terminal de la enzima lítica del fago Tuc2009 de *Lactococcus lactis* y el C-terminal de LytA (Sheehan y cols., 1996) proporcionan nuevos argumentos experimentales a la teoría de la evolución modular de las proteínas que propone que las proteínas de aparición reciente han evolucionado a partir de la unión de unidades polipeptídicas preexistentes (Campbell, 1988).

5. OBJETIVOS

La biología molecular del sistema de neumococo y sus bacteriófagos es poco conocida a pesar de ser el microorganismo que sirvió para demostrar que el DNA era la macromolécula responsable de la herencia de los caracteres, un hecho que marcó el inicio de la genética molecular (Griffith, 1928; Avery y cols., 1944). Las dificultades que se presentan en el laboratorio para multiplicar a neumococo en medios simples de cultivo ha supuesto un problema insalvable a la hora de desarrollar estudios genéticos en este sistema y, asimismo, ha impedido poner a punto técnicas de transferencia genética que implicaran el uso de partículas fágicas. El empleo de los antibióticos β -lactámicos representó un punto de inflexión para frenar los efectos devastadores causados por neumococo hasta los años 40. No obstante, en la actualidad, este microorganismo ha vuelto a producir una gran mortalidad en el mundo (Sahn, 1990). En consecuencia, es muy importante potenciar investigaciones que contribuyan a desarrollar nuevos instrumentos para combatir los efectos patógenos causados por *S. pneumoniae*. En general, el estudio de los fagos ha constituido una eficaz herramienta de trabajo para conseguir un mejor conocimiento de la bacteria huésped y, en este sentido, cabe destacar la reciente sugerencia de que los fagos podrían jugar un importante papel en la evolución y transferencia de los determinantes de virulencia en bacterias (Cheetham y Katz, 1995). Finalmente, existen muy pocos ejemplos en la literatura sobre la organización funcional de fagos que replican de forma linear. El bacteriófago ϕ 29 y los fagos con él relacionados, en bacterias Gram-positivas, y el fago PRD1 en Gram-negativas, proporcionan los modelos experimentales sobre este peculiar mecanismo de replicación de DNA en los sistemas procarióticos (Salas, 1991).

Por todo ello, nos pareció interesante estudiar las características moleculares del bacteriófago Cp-1 que, por su estabilidad y pequeño tamaño, le confieren una gran versatilidad como herramienta de trabajo para conseguir una detallada información acerca de los productos génicos que codifica y de las funciones fisiológicas que desempeñan tales productos.

Teniendo en cuenta todo lo citado anteriormente, nos propusimos en esta Memoria los siguientes objetivos experimentales:

- 1.- Secuenciación completa del genoma del bacteriófago Cp-1.
- 2.- Elaboración del mapa transcripcional del fago.

3.- Estudio de algunas de las proteínas codificadas por el fago, en particular las implicadas en los procesos líticos que permiten la liberación de la progenie.

4.- Estudio del mecanismo de iniciación de la replicación.

II. MATERIALES Y MÉTODOS

1. CEPAS BACTERIANAS, BACTERIOFAGOS Y PLÁSMIDOS

Para el desarrollo de este trabajo se utilizaron las cepas bacterianas, bacteriófagos y plásmidos que se detallan en la Tabla 4.

Tabla 4. Cepas bacterianas, bacteriófagos y plásmidos

Cepa, fago o plásmido	Genotipo o fenotipo relevante	Referencia/Fuente
<i>E. coli</i>		
TG1	F' (<i>lacI^h lacZΔM15</i>)	Gibson, 1984
DH1	<i>supE44 hsdR17 recA1 endA1</i>	Sambrook y cols., 1989
DH5α	(<i>Φ80lacZΔM15</i>) <i>endA1 recA1 hsdR17(r_K⁻ m_K⁻) supE44 ΔlacU169</i>	Hanahan, 1983
HB101	<i>supE 44 hsdR 20 (r_B⁻ m_B⁻) recA13</i>	Boyer y Roulland-Dussoix, 1969
C600	<i>thi-1, leuB, thr-1</i>	Appleyard, 1954
CM12	MC4100 <i>recA⁻</i>	Sancar y cols., 1979
K514	<i>supE44 thr1 leu6 (r_K⁻ m_K⁺)</i>	Laboratorio M. Salas
BL21(DE3)	F' <i>hsdS (r_B⁻ m_B⁻)</i> , polimerasa del fago T7 bajo el control del promotor <i>lac UV54</i>	Studier y Moffatt, 1986
LE392	F' <i>hsdR574 (r_K⁻ m_K⁻) supE44 supF58</i>	Sambrook y cols., 1989
<i>S. pneumoniae</i>		
R6	cepa salvaje	Rockefeller Univ.
R6st	Sm ^R	Rockefeller Univ.
M31	Hex ⁻ <i>ΔlytA</i>	Sánchez-Puelles y cols., 1986a
Fagos		
Cp-1		Ronda y cols., 1981
M13tg131		Sambrook y cols., 1989
Lambda <i>Sam7cI857</i>		Promega
Plásmidos		
pUC18	Ap ^R	Vieira y Messing, 1982
pUC19	Ap ^R	Yanisch-Perron y cols., 1985
pBR322	Ap ^R , Cm ^R	Bolívar y cols., 1977
pBR325	Ap ^R , Cm ^R , Tc ^R	Bolívar, 1978
pT7-4	Ap ^R	Tabor y Richardson, 1985
pT7-7	Ap ^R	Tabor y Richardson, 1985
pIN-III-A3	Ap ^R	Inouye e Inouye, 1985
pLSE1	Ery ^R , Tc ^R	Ronda y cols., 1988
pLysS	Cm ^R	Studier, 1991
pNM185	Km ^R <i>xylS</i>	Mermod y cols., 1986
pGL80	Ap ^R <i>lytA</i>	García y cols., 1986a
pCIP50	Ap ^R <i>cpl1</i> (derivado de pBR325)	García y cols., 1987
pCIP100	Ap ^R <i>cpl1</i> (derivado de pIN-III-A3)	Sanz y García, 1990

2. TITULACIÓN, MULTIPLICACIÓN Y PURIFICACIÓN DE FAGOS

La titulación de Cp-1 se llevó a cabo utilizando como bacteria huésped la cepa R6 de *S. pneumoniae* cultivada inicialmente en medio semisintético (CpH8) (Tomasz, 1970) a 37°C sin agitación para pasarlo a medio sólido KCAT conteniendo catalasa en el recuento de las placas de lisis. Los títulos obtenidos fueron aproximadamente 10^8 UFP/ml para lisados en cultivos líquidos y 10^{10} - 10^{11} UFP/ml en muestras purificadas mediante gradientes de CsCl.

La obtención del fago se realizó incubando a 37°C, sin agitación, la cepa R6st de *S. pneumoniae* en medio CpH8 enriquecido con extracto de levadura (0,08%) hasta alcanzar una densidad celular de 6×10^8 UFC/ml. A continuación se hacía una dilución 1:10 en medio M3 (triptona 17 g/l, glucosa 2,5 g/l, NaCl 5 g/l, neopeptona 5 g/l y tampón fosfato 50 mM pH 8,0) precalentado. Se incubó hasta una densidad celular de $8,3 \times 10^7$ UFC/ml y se infectó con el fago a una m.d.i. de 1/3, y se pasó el cultivo a 30°C que es la temperatura adecuada para la multiplicación del fago. La incubación se continuó hasta que se produjo la lisis, se eliminó el resto de las células no lisadas mediante centrifugación y se recogió el sobrenadante donde se encuentran las partículas fágicas. Se precipitaron éstas con PEG 6000 (10% p/v) a 4°C toda la noche. Se resuspendieron en TBT (Tris 12,1 g/l, NaCl 5,8 g/l, $Mg_2Cl_2 \cdot 6H_2O$ 2 g/l, HCl fumante 5ml/l) y se purificó el fago usando una centrifugación, en dos pasos, en gradiente de CsCl (Ronda y cols., 1981). El fago obtenido de la segunda centrifugación, que se detecta formando una banda azul, se dializó frente a TBT y se congeló rápidamente en nieve carbónica y etanol a -70°C en 10% de glicerol. A partir de 3 l de medio M3 se obtuvieron 5 ó 6 ml de fago del título anteriormente señalado.

Para la titulación del fago lambda se incubó la cepa de *E. coli* correspondiente a 30°C en medio LB suplementado con maltosa 0,2%, $MgSO_4$ 10 mM. Cuando los cultivos alcanzaron la DO_{600} de alrededor de 0,8, se recogieron muestras de 200 μ l y se incubaron a 37°C durante 30 min con diferentes diluciones de una preparación del fago lambda. A continuación se añadieron 5 ml de LB-agar 0,8% con $MgSO_4$ 10 mM y se contaron las placas de lisis en LB después de una incubación a 30 o 37°C, según la cepa, durante toda la noche.

Para la titulación del fago M13tg131 se siguieron las instrucciones descritas por Sambrook y cols. (1989).

3. PREPARACIÓN DEL DNA FÁGICO

El DNA de Cp-1 se obtuvo por extracción fenólica (v/v) después del tratamiento del fago con proteinasa K (50 µg/ml) en un tampón Tris-HCl 10 mM pH 8,0, EDTA 10 mM, NaCl 10 mM y SDS 0,5% a 37°C durante 3 ó 4 horas. Este tratamiento evita que el complejo DNA-proteína que se aísla del virión quede retenido en la interfase (López y cols., 1984). Después del tratamiento con fenol, el DNA se precipitó y se disolvió en TE.

4. MEDIOS Y CONDICIONES DE CULTIVO

Las cepas se conservaron congeladas a -70 °C en medio de cultivo al que se añadió glicerol al 20 y al 10% (v/v) para el caso de *E. coli* y *S. pneumoniae*, respectivamente. En el momento de sembrarlas se descongelaron, incubándose en los medios correspondientes a 37°C, a menos que se indique lo contrario, con agitación para las cepas de *E. coli* y sin agitación para las cepas de *S. pneumoniae*.

4.1. *E. coli*

Para el cultivo de *E. coli* se utilizó el medio LB descrito por Sambrook y cols. (1989), añadiendo 1,5% agar (p/v) para cultivos en medio sólido. El medio M9 suplementado con tiamina 2 µg/ml, MgSO₄ 2 µg/ml, CaCl₂ 0,1 µg/ml y glucosa 0,2% (p/v) se usó para el estudio de la expresión de genes bajo el control del promotor $\phi 10$ del bacteriófago T7. La concentración de antibiótico añadida al medio de cultivo de cepas resistentes fue de 100 µg/ml para la ampicilina (Ap), 10 µg/ml para la tetraciclina (Tc), 30-60 µg/ml para el cloranfenicol (Cm) y 50 µg/ml para la kanamicina (Km).

El crecimiento de las cepas de *E. coli* se siguió por turbidimetría a 600 nm con un espectrofotómetro Shimadzu UV-260.

4.2. *S. pneumoniae*

Para el cultivo de *S. pneumoniae* se utilizó CpH8 suplementado con extracto de levadura (0,08%) (Tomasz, 1970). Para cultivos en medio sólido se añadió al medio

anterior, agar 1% (p/v) y catalasa (250 U/ml) (Boehringer Mannheim) aunque en otros casos se utilizaron placas de agar sangre [Tryptic-Soy agar (Difco), suplementado con 5% de sangre desfibrinada de carnero].

Las cepas transformantes de neumococo que contienen derivados de pLSE1 (Ronda y cols., 1988) se seleccionaron en un medio con Tc a 0,8 µg/ml (medio líquido) o 1 µg/ml (medio sólido) ó bien eritromicina (Ery) (10 µg/ml) o lincomicina (Lc) (1 µg/ml).

La multiplicación de este microorganismo se siguió por nefelometría con el empleo de un nefelómetro Coleman modelo 9, calibrado para recuento de viables en neumococo donde 1 N= $7,5 \times 10^5$ UFC/ml.

5. PROCEDIMIENTOS DE TRANSFORMACIÓN GENÉTICA

5.1. *E. coli*

Las cepas de *E. coli* se transformaron utilizando el método del RbCl (Kushner, 1978) o bien mediante electroporación (Dower y cols., 1988) utilizando un equipo Gene Pulser de BioRad.

5.2. *S. pneumoniae*

La preparación de cepas competentes de neumococo ha sido descrito por Tomasz (1970). Los cultivos de neumococo poseen competencia natural que alcanza sus valores más altos en una determinada etapa del crecimiento (la comprendida entre $2,1 \times 10^8$ y $2,25 \times 10^8$ UFC/ml). Una vez que el cultivo alcanza esta concentración celular, se le añade glicerol 10% (v/v) y se congela rápidamente. Las células conservadas a -70°C mantienen la competencia, sin pérdida apreciable, durante largos periodos de tiempo.

El método de transformación utilizado fue esencialmente el descrito por Barany y Tomasz (1980), excepto que la incubación con el DNA transformante se realizó a 30°C durante 30 min en medio CpH8 precalentado suplementado con seroalbúmina bovina (BSA) 0,08% (p/v). Después los cultivos se incubaron a 37°C durante 90 min en medio precalentado CpH8 con extracto de levadura 0,08% y con BSA 0,08% (fase de expresión), antes de la siembra a 37°C en el medio sólido adecuado.

6. ELECTROFORESIS DE DNA EN GELES DE AGAROSA

Se utilizaron geles de agarosa al 0,7 ó 1% en tampón TAE (Tris-HCl 40 mM, ácido acético 20 mM, EDTA 2 mM, pH 8,1), utilizando el mismo tampón como electrolito. A las muestras se les añadió 1-1,5/10 de su volumen de una solución compuesta por Ficoll 400 30% (p/v), azul de bromofenol 0,2% (p/v), xilencianol 0,2% y EDTA 40 mM pH 8,0. La electroforesis se realizó a 100-150 V durante 60-90 min y, una vez finalizada, los geles se tiñeron con bromuro de etidio (BrEt) (3 µg/ml) para su observación mediante irradiación UV. Como marcadores de tamaño, se utilizaron el DNA del fago lambda digerido con la enzima de restricción *Bst*EII y la forma replicativa del DNA del fago ϕ X174 digerido con *Hae*III.

7. AISLAMIENTO Y PURIFICACIÓN DE FRAGMENTOS DE DNA

Los fragmentos resultantes de la digestión del DNA con las enzimas de restricción o los obtenidos mediante PCR, se aislaron y purificaron por tres procedimientos distintos.

7.1. Técnica de Gene-Clean

Los fragmentos que se desean aislar se separaron primero mediante una electroforesis en geles de agarosa. Posteriormente, se recortaron las bandas de interés y el DNA se purificó de la agarosa mediante su adsorción a diminutas partículas de vidrio siguiendo la técnica de Gene-Clean, tal y como recomiendan los fabricantes (Bio 101 Inc., La Jolla, CA, USA).

7.2. Técnica de la β -agarasa

El empleo de esta técnica también requirió la separación previa de los fragmentos mediante electroforesis en geles de agarosa de bajo punto de fusión y posterior tratamiento de la banda extraída con la β -agarasa, que hidroliza la agarosa, liberando el DNA. Para ello fue necesario fundir previamente la banda de agarosa recortada en el tampón adecuado (Tris-HCl 10 mM, pH 6,5; EDTA 10 mM) antes del tratamiento con β -agarasa durante una

hora a 40-42°C. Tras tratar con fenol y precipitar el DNA, éste se disolvió en el volumen adecuado de TE. Tanto la enzima como el tampón se obtuvieron de BioLabs (New England, USA).

7.3. Geles de agarosa de bajo punto de fusión

Para algunos aislamientos de fragmentos de DNA se utilizaron geles de agarosa de bajo punto de fusión (BioRad) de 0,7-1% en tampón TAE. Una vez realizada la electroforesis, se recortó del gel la banda deseada en cada caso y se resuspendió en tampón TE (Tris-HCl 10 mM pH 8,0; EDTA 1mM) con NaCl 0,25 M. La agarosa se fundió calentándola a 65°C durante 5 minutos. A continuación, la muestra se trató con fenol y posteriormente con éter etílico (v/v), y se incubó después a 65°C durante 10-15 min con el fin de eliminar los posibles restos de éter. El DNA se precipitó con dos volúmenes y medio de etanol absoluto durante 45 min a -70°C. Tras una centrifugación a $10.000 \times g$ durante 10 min, el precipitado se lavó con etanol 70% y finalmente se disolvió en un volumen adecuado de tampón TE.

8. TÉCNICA DE LA REACCIÓN EN CADENA DE LA POLIMERASA (PCR)

Las amplificaciones se realizaron en el aparato Pharmacia LKB, Gene ATAQ Controller utilizando, como molde, el DNA entero de Cp-1 y, como cebadores para la DNA polimerasa, oligonucleótidos sintéticos:

a) la amplificación del fragmento del genoma comprendido entre los nt 18.074 y 18.954 se llevó a cabo utilizando la DNA polimerasa *PfuI* (Stratagene). El programa empleado consistió en 30 ciclos de desnaturalización, los cinco primeros a 98°C y el resto a 95°C durante 1 min, seguido de hibridación a 55°C de los oligonucleótidos al molde durante 2 min y la reacción de polimerización a 72°C por la enzima. El fragmento amplificado obtenido se trató con fenol, se precipitó disolviéndose en TE, purificándose después por la técnica de la β -agarosa. Los extremos del fragmento amplificado fueron hechos romos con el fragmento klenow de la DNA polimerasa I de *E. coli* (klenow) y el DNA se ligó a la forma replicativa del DNA del fago M13tg131 cortado con *HincII* y a pUC18 digerido con *SmaI* y tratado con fosfatasa alcalina.

b) el resto de las amplificaciones se llevaron a cabo con la DNA polimerasa termoestable Replitem™ (Epicentre). El programa consistió en 30 ciclos de desnaturalización, los 7 primeros a 98°C y el resto a 95°C, durante 1 min, seguido de hibridación a 50°C de los oligonucleótidos al molde durante 2 min y la reacción de polimerización a 72°C durante 2,5 min llevada a cabo por la enzima, excepto para la obtención de la sonda 3 utilizada en los ensayos de "Northern-blot", en el que la hibridación fue a 45°C y la reacción de polimerización duró 2 min. Los fragmentos amplificados se extrajeron con fenol, se precipitaron disolviéndose en TE, purificándose después por las técnicas descritas en el apartado 7.

9. MANIPULACIÓN DEL DNA CON ENZIMAS DE USO COMÚN EN BIOLOGÍA MOLECULAR

Las endonucleasas de restricción se obtuvieron de Boehringer Mannheim, Amersham, Pharmacia y New England Biolabs. La fosfatasa alcalina de intestino de ternera se obtuvo de Boehringer Mannheim. Amersham suministró la DNA ligasa, la nucleasa BAL31 y la polinucleótido quinasa del fago T4, así como Klenow. La transcriptasa inversa fue suministrada por Promega. Las enzimas y sus tampones correspondientes se utilizaron siguiendo las indicaciones recomendadas por las casas suministradoras.

Los fragmentos de restricción purificados *Hind*III (1.449-10.736 nt), *Hinc*II (9.330-13.438 nt) y *Sau*96A-*Sph*I (12.821-16.754 nt) se digirieron con la nucleasa BAL31 incubándose a 30°C. A diferentes tiempos se sacó una alícuota de la mezcla y se paró la reacción con EGTA 20 mM a 4°C con el fin de obtener una serie de fragmentos de distintos tamaños. Estos fragmentos se trataron con fenol, se precipitaron y se disolvieron en TE. Posteriormente se rellenaron con Klenow para ligarlos a pUC18 digerido previamente con *Sma*I y tratado con fosfatasa alcalina (Sambrook y cols., 1989).

Para obtener los plásmidos recombinantes con insertos de la parte izquierda del genoma de Cp-1, se digirió el DNA del fago con *Hind*III, y se purificó el fragmento de aproximadamente 1,45 kb por Gene-Clean. Parte del fragmento obtenido se digirió durante 15 min a 37°C con *Ssp*I y, después de inactivar la digestión, se ligó la mezcla a pUC18 digerido doblemente con *Sma*I y *Hind*III. El resto del producto del Gene-Clean se digirió durante 30 min a 37°C y se ligó a pUC18 cortado con *Sma*I y tratado con fosfatasa alcalina.

10. TÉCNICAS DE HIBRIDACIÓN DEL DNA Y RNA

Los fragmentos de DNA marcados radiactivamente (sondas) se obtuvieron mediante la técnica del “random-primer” utilizando $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ (400 Ci/mmol, 10 $\mu\text{Ci}/\mu\text{l}$) (Amersham), Klenow, la solución de hexanucleótidos que actúan de iniciadores (Pharmacia) y los dNTPs, según se indica en Sambrook y cols. (1989). Todas las hibridaciones y lavados se hicieron a 65° C.

El marcaje de oligonucleótidos sintéticos se llevo a cabo mediante la utilización de la polinucleótido quinasa T4 utilizando $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (5.000 Ci/mmol, 10 $\mu\text{Ci}/\mu\text{l}$) siguiendo las indicaciones de Sambrook y cols. (1989). Las hibridaciones se realizaron a 30°C y los lavados se hicieron a temperatura ambiente.

10.1. Técnica de “Southern-blot”

La hibridación del DNA mediante la técnica descrita por Southern (1975), se realizó siguiendo el protocolo de Sambrook y cols. (1989). Las membranas de nilón para la transferencia del DNA se obtuvieron de Schleicher and Schuell. Las bandas radiactivas se detectaron utilizando películas HyperfilmTM-MP (Amersham) y pantallas amplificadoras Dupont, Cronex Lightning Plus, a -70°C. La deshibridación de las membranas se realizó con un primer tratamiento con NaOH 0,4 M durante 30 min a 45°C, y un segundo tratamiento con 0,1 \times SSC (NaCl 0,15 mM, citrato sódico 0,15 mM) y SDS 0,1% otros 30 min a la misma temperatura.

10.2. Hibridación de DNA sobre filtros de colonias celulares

Las colonias de *E. coli* se transfirieron a filtros de nitrocelulosa (Millipore HATF, tamaño de poro 0,45 μm) como se indica en Sambrook y cols. (1989). Las células se lisaron sobre el filtro mediante tratamientos sucesivos, de 5 min de duración cada uno, con las soluciones que se indican: i) NaCl 1,5 M y NaOH 0,5 M (2 veces); ii) Tris-HCl 1M, pH 7,0; iii) Tris-HCl 0,5 M pH 7,0 y NaCl 1,5 M; iv) 2 \times SSC (NaCl 0,3 M, citrato sódico 0,003 M). La fijación del DNA al filtro se realizó mediante incubación a 80°C durante 2 horas. Para eliminar los restos celulares, los filtros se hirvieron en agua destilada durante 10 min.

La hibridación con la sonda radiactiva se realizó como se describe en Sambrook y cols. (1989).

10.3. Técnica de "Dot-blot"

El DNA se desnaturalizó previamente con NaOH a una concentración final de 0,5 M, durante 10 min a 37°C. A continuación, se colocó la membrana de nilón entre las dos placas del aparato, se hizo el vacío y se añadió el DNA en cada pocillo. Una vez que el DNA quedó retenido en la membrana, se desnaturalizó 1 min con una solución 0,5 M NaOH, NaCl 1,5 M, se neutralizó en Tris-HCl 0,5 M pH 7,0, NaCl 1,5 M, durante 1 min y se irradió con UV para fijar el DNA a la membrana. Después se hibridó la membrana de igual manera que en "Southern-blot".

10.4. Técnica de "Northern-blot"

Tras el aislamiento de RNA, tal como se indica en el apartado 13, las muestras se cargaron en geles de agarosa 1-2% con formaldehído y se sometieron a electroforesis a 9 V/cm durante la noche. El gel se dializó varias veces contra agua y se trató 25 min con NaOH 50 mM, NaCl 0,1 M, después 25 min con Tris-HCl 0,1 M pH 7,5 para neutralizar y, por último, 25 min con 2 × SSC. La transferencia de RNA se realizó mediante vacío utilizando como tampón 20 × SSC. La membrana se secó durante 30 min a 65°C y, posteriormente, se fijó el RNA mediante irradiación UV durante 10 min.

11. PREPARACIÓN DE PLÁSMIDOS

La preparación de plásmidos a partir de cepas de *E. coli* se realizó utilizando el método rápido de lisis alcalina (Birnboim y Doly, 1979) y el método de purificación de DNA por centrifugación en gradiente de CsCl (Sambrook y cols., 1989).

La preparación de plásmidos a partir de cultivos de neumococo en fase exponencial de crecimiento se llevó a cabo siguiendo el método de Birnboim y Doly (1979), con la modificación consistente en que la lisis celular se realizó mediante la resuspensión de las células en tampón fosfato sódico 50 mM, pH 8,0, y posterior incubación a 37°C durante 15

min en presencia de DOC (Fluka) (1% p/v). Para la extracción de plásmidos de cepas con fenotipo *LytA*⁻, las células se trataron con amidasa *LytA* durante 30 min a 37°C antes de la adición de DOC.

12. SECUENCIACIÓN DE DNA

El método seguido para la secuenciación de DNA fue el descrito por Sanger y cols., (1977), utilizando el T7 Sequencing™ Kit (Pharmacia) y [α -³⁵S]dATP (Amersham), siguiendo las recomendaciones de los fabricantes y cubetas de Lagoplast (36 × 50 cm) o BioRad (38 × 52 cm). Para preparar la cadena sencilla del DNA plasmídico para las reacciones de secuencia, los plásmidos se desnaturalizaron con NaOH 0,2 M durante 5-8 min a temperatura ambiente, se neutralizaron con acetato amónico 0,2 M y el DNA se precipitó con etanol. Para secuenciar el DNA de Cp-1 intacto, éste fue desnaturalizado con NaOH 0,4 M durante 18 min a temperatura ambiente. El resto del proceso es igual al descrito para los plásmidos. Para eliminar las compresiones y ambigüedades de la secuencia se usó el kit de Pharmacia Deaza G/A T7 Sequencing™.

12.1. Análisis de la secuencia de DNA

Para el análisis de la secuencia de DNA se utilizó el paquete de programas para Biología Molecular de la Universidad de Wisconsin (WINP) denominado GCG (Genetics Computer Group) (Devereux y cols., 1984). presente en el VAX del Centro Nacional de Biotecnología de Madrid. Los bancos de datos a los que se ha tenido acceso fueron: GenBank/EMBL, Swissprot, PIR, PROSITE; y los programas utilizados fueron: FETCH, FASTA, TFASTA, BESTFIT, PILEUP, MOTIFS y PEPLOT. Para la predicción de la estructura secundaria de las proteínas codificadas por las ORFs además del PEPLOT, se utilizó el programa del EMBL, Prediction Protein del grupo de Sander.

13. AISLAMIENTO DE RNA

La cepa R6st de *S. pneumoniae* se incubó a 37°C en medio CpH8 suplementado con extracto de levadura 0.08% hasta una densidad celular de $7,5 \times 10^8$ UFC/ml. El cultivo se

infectó con Cp-1 a una m.d.i. de 8 y se incubó a 30°C; el cultivo control sin infectar se procesó de la misma manera. Se tomaron muestras a diferentes tiempos, se centrifugaron y las células se lavaron con TE (10:10) (Tris-HCl 10 mM: EDTA 10 mM). A continuación, las células se lisaron con TE (10:10) conteniendo DOC 0,1%, congelándose y descongelándose dos veces sucesivas en N₂ líquido. Posteriormente, las muestras se incubaron con SDS 0,5%, acetato sódico 0,1% pH 4,0 y fenol ácido equilibrado con agua durante 5 min a 60°C. Las muestras se centrifugaron a 4°C recuperándose la fase acuosa y extrayéndose el RNA con fenol caliente y posterior precipitación con etanol absoluto. El RNA extraído se disolvió en el volumen adecuado de TE (10:10) y se conservó a -70°C. Para la obtención de RNA de cultivos tratados con Cm, se siguió el mismo proceso descrito anteriormente, añadiendo Cm (100 µg/ml concentración final) al cultivo de R6st 5 min antes de la infección.

La concentración de RNA se estimó a partir de la medida de absorbancia a 260 nm, considerando que una unidad de absorbancia en cubetas de 1 cm de paso óptico equivale a una concentración de 40 µg/ml. Para comprobar la presencia de las bandas esperadas de RNA, las muestras se analizaron en un gel de agarosa al 1%.

14. DETERMINACIÓN DE LOS SITIOS DE INICIACIÓN DE LA TRANSCRIPCIÓN MEDIANTE EL EMPLEO DE LA TÉCNICA DE “PRIMER EXTENSION”

Para localizar los diferentes sitios de iniciación de la transcripción del genoma del bacteriófago Cp-1 se siguieron dos procedimientos distintos, ambos basados en la técnica de extensión del cDNA a partir de un iniciador o “primer extension”:

a) Marcando el oligonucleótido con [γ -³²P]ATP, según el protocolo descrito por Sambrook y cols. (1989). Los oligonucleótidos (7 pmol) se marcaron con 50 µCi de [γ -³²P]ATP (5.000 Ci/mmol, 10 µCi/µl). Posteriormente se mezclaron 1 pmol de oligonucleótido y 19 µg de RNA precipitando la mezcla según el procedimiento descrito por Sambrook y cols. (1989). El precipitado se resuspendió en 30 µl de tampón de hibridación (PIPES 40 mM pH 6,4, EDTA 1 mM pH 8,0, NaCl 0,4 M y formamida 50%), desnaturalizando a 85°C durante 10 min e incubando la mezcla 12 horas a 30°C para permitir la hibridación. El híbrido RNA-oligonucleótido se precipitó con 2,5 volúmenes de etanol absoluto

resuspendiéndolo posteriormente en la mezcla de reacción de la transcriptasa inversa AMV (1 × tampón AMV, dNTPs 0,9 mM pH 7,0, RNAsina 0,4 U, y transcriptasa inversa AMV 3 U). La reacción de extensión transcurrió durante una hora a 42°C y se detuvo con NaOH a una concentración final de 0,4 M, manteniéndose 12 horas a temperatura ambiente para desnaturalizar los ácidos nucleicos. El siguiente paso consistió en la neutralización de la mezcla con ácido acético 0,4 M y posterior precipitación con etanol, resuspendiendo el precipitado en una solución de colorante (azul de bromofenol 0,15%, xilencianol 0,15%, EDTA 5 mM pH 7,5 y formamida desionizada 48,75%). Para analizar la longitud de los productos de la extensión y, por tanto, el punto de iniciación de la transcripción, se realizó una electroforesis en paralelo cargando las muestras en un gel del 6% poliacrilamida conteniendo 8 M urea, detectándose el resultado por autorradiografía. Como control se cargaron los plásmidos recombinantes, o el DNA de Cp-1 intacto, secuenciados con los mismos oligonucleótidos de cada reacción de “primer extension”.

a) Marcando el DNA complementario del RNA con [α -³²P]dCTP, según el método descrito por Chandry y cols. (1994). Para ello, se calentaron a 90°C durante 1 min, 3 μ l de la mezcla que contenía 0,25 μ g de oligonucleótido y 20 μ g de RNA y posteriormente se incubó la mezcla a 42°C durante 5 min para realizar el anillamiento RNA-oligonucleótido. El cDNA es sintetizado en un volumen de reacción de 10 μ l que contenía 24 U de la enzima transcriptasa inversa (AMV), 15 μ Ci de [α -³²P]dCTP, dCTP 10 μ M, dATP 100 μ M, dGTP 100 μ M, dTTP 100 μ M, Tris-HCl 50 mM pH 8,5, KCl 50 mM, MgCl₂ 20 mM y DTT 10 mM junto con la mezcla de anillamiento, RNA-oligonucleótido. La reacción se incubó a 42°C durante 30 min y se paró con 5 μ l de formamida 0,5%, EDTA 20 mM, azul de bromofenol 0,05% y xilencianol 0,05%. El resultado se analizó como en el método anterior.

Este procedimiento es mucho más rápido que el anterior y tiene la ventaja de conseguir una mejor desnaturalización cuando se trabaja con material genético de elevado contenido en estructura secundaria.

15. ELECTROFORESIS DE PROTEÍNAS EN GELES DE POLIACRILAMIDA-SDS

Las electroforesis analíticas de proteínas se realizaron en todos los casos en condiciones desnaturalizantes, en geles de poliacrilamida-SDS, según la técnica descrita por

Laemmli (1970) el porcentaje de acrilamida utilizado varió entre 8 y 17,5%. Las muestras se hirvieron durante 5 min en tampón de ruptura (Tris-HCl 62,5 mM pH 6,8; SDS 2%, β -mercaptoetanol 5%, glicerol 10% y azul de bromofenol 0,005%). Las electroforesis se realizaron a temperatura ambiente y corriente constante (50 mA), utilizando un electrolito que contenía Tris-HCl 0,025 M, glicina 0,192 M y SDS 0,1%; se usaron placas de 160 \times 100 mm con espaciadores de 2 mm. Las proteínas de los geles se tiñeron con azul de Coomassie R-250, según se describe en Swank y Munkress (1971). Las proteínas empleadas como marcadores de peso molecular eran de BioRad: fosforilasa b (92,5 kDa); BSA (66,2 kDa); ovoalbúmina (43 kDa); anhidrasa carbónica (31 kDa); inhibidor de tripsina (21,5 kDa) y lisozima (14,4 kDa).

16. OBTENCIÓN DE EXTRACTOS DE CULTIVOS CELULARES

16.1. Extractos preparados de cepas de *E. coli* y *S. pneumoniae*

Para obtener extractos de *E. coli* y *S. pneumoniae*, se cultivaron estos microorganismos en 10 ml del medio indicado en cada caso, a 37°C (con y sin inductor cuando sea preciso) hasta alcanzar la mitad de la fase exponencial. Los cultivos se enfriaron a 4°C y se centrifugaron a 6000 \times g durante 10 min, resuspendiéndose las células en 500 μ l de tampón fosfato sódico 20 mM pH 6,9, EDTA 1 mM, (en algunas ocasiones en tampón de ruptura). Una vez homogeneizada la suspensión, las células se rompieron con ultrasonidos utilizando un aparato modelo B-12A de Branson Sonic Power co. (Danbury, Connecticut) mediante cuatro tratamientos de 10 s cada uno con intervalos de 1 min, manteniéndose siempre la muestra a 4°C. La suspensión así tratada se centrifugó a 4°C y 10.000 \times g durante 10 min para eliminar los restos celulares. El precipitado se resuspendió en el mismo tampón fosfato y los extractos se conservaron a -20°C. La fracción proteica se valoró mediante el método de Bradford (1976) y se analizó mediante geles de poliacrilamida-SDS. Cuando se partió de volúmenes mayores, se centrifugaron los cultivos 20 min a 6.000 \times g y los precipitados se resuspendieron en tampón fosfato 20 mM pH 6,9, EDTA 1 mM, las células se rompieron mediante una prensa de French (American Instruments Company) a una presión de 7,6 atm. Se eliminaron los restos de DNA de los extractos con un tratamiento con polietilenimina (0,3% para DO₂₆₀ de 120) en hielo durante

una hora. Después se centrifugaron 15 min a $12.000 \times g$ y se recogieron los sobrenadantes. Se precipitaron las proteínas con sulfato amónico al 65%, se resuspendieron en 1 ml de 20 mM tampón fosfato pH 6,9, EDTA 1 mM y se dializaron frente a tampón fosfato 20 mM pH 6,9 para eliminar el sulfato amónico.

Para la obtención de extractos de cepas que contienen como plásmidos pT7-7 o algún derivado, se cultivaron en medio M9 suplementado como se indica en el apartado 4.1 a 37°C hasta alcanzar una DO_{600} entre 0,7-1, entonces se indujo con IPTG a diferentes concentraciones (0,01, 0,1, 0,4 y 1 mM). A continuación, se recogieron a distintos tiempos después de la inducción muestras de 1,5 ml, que se enfriaron a 4°C y se centrifugaron a $10.000 \times g$ durante 5 min, resuspendiéndose las células en 300 μ l de tampón fosfato sódico 20 mM pH 6,9, EDTA 2,5 mM. Una vez homogeneizada la suspensión, las células se rompieron como se ha descrito anteriormente y se siguió el protocolo descrito anteriormente. Como controles de estos experimentos se obtuvieron extractos de las cepas recombinantes sin inducir y de la cepa que contiene pT7-7, con y sin inducción.

La cepa *E. coli* DH5 α (pAMR61) se incubó a 37°C en 10 ml de medio LB + Ap (100 μ g/ml) sin IPTG, puesto que el gen de la proteína de la cabeza de Cp-1 se expresó con los mismos niveles sin inducir que inducido, hasta la fase estacionaria. Se centrifugaron las células y se resuspendieron en 1,5 ml de TSE (Tris-HCl 10 mM pH 7,0, EDTA 0,5 mM). Se volvieron a centrifugar y se resuspendieron en 500 μ l de TSE. Se centrifugaron las muestras tratadas con ultrasonidos 10 s dos veces con intervalos de 1 min. Se centrifugaron los sonicados y se recogió el sobrenadante, y el precipitado se resuspendió en 500 μ l de TSE. Al cargar estos extractos en un gel del 10% de poliacrilamida-SDS se observó que la proteína estaba tanto en forma soluble como en insoluble. El extracto con la proteína soluble expresada en *E. coli* es el que se utilizó para determinar la secuencia N-terminal de la proteína mayoritaria de la cabeza de Cp-1. Para obtener la fracción insoluble enriquecida en la proteína, se siguió el proceso con la fracción del precipitado resuspendida en TSE. Se centrifugó, y se resuspendió en 1 ml de urea 3 M, Triton X-100 1% en TSE. La mezcla se agitó a temperatura ambiente durante 30 min. Este proceso se repitió tres veces y para terminar se centrifugó el extracto 10 min y se lavó tres veces con TSE. Al final se resuspendió en 500 μ l de TSE.

Para la obtención de extractos de M31(pAMR71) infectados con Dp-1, se incubó esta cepa en 100 ml de Cph8 suplementado con extracto de levadura 0.08% más Lc y, a

una densidad celular de $7,5 \times 10^7$ UFC/ml, se infectó con el fago Dp-1 a una m.d.i. de 1. Cuando el cultivo se empezó a lisar se recogió y se procesó como se ha indicado previamente.

Para la obtención de extractos de M31(pAMR71) tratados con UV, previamente se establecieron las condiciones adecuadas de irradiación con UV. Posteriormente, 100 ml del cultivo se irradiaron a una densidad celular de $1,9 \times 10^8$ UFC/ml durante 45 s, y se incubaron en oscuridad hasta llegar a una densidad celular de $4,8 \times 10^8$ UFC/ml. A continuación, se centrifugaron y se procedió de forma similar a la descrita previamente.

Para la obtención de extractos de M31(pAMR71) tratados con mitomicina (Boehringer Mannheim), se probaron distintas concentraciones de esta droga y se concluyó que la adecuada era 20 ng/ml añadida al cultivo cuando está a una densidad de $1,2 \times 10^8$ UFC/ml. La obtención de los extractos fue la misma que en los casos anteriores.

16.2. Extractos de neumococo infectados con Cp-1 para realizar ensayos de iniciación de la replicación

La cepa de R6st de *S. pneumoniae* se infectó con Cp-1 a una m.d.i. de 2,5 y se incubó a 30°C durante 90 min. Las células se recogieron y se lavaron con Tris-HCl 50 mM pH 7,5. Las células se lisaron con alúmina durante 15 min a 4°C y después se centrifugaron para eliminar la alúmina y los restos celulares. El sobrenadante se trató con polietilenimina 0,3% por cada 120 U de DO₂₆₀ para eliminar el DNA bacteriano y el complejo TP-DNA del fago. Tras la centrifugación durante 10 min a $10.000 \times g$ las proteínas presentes en el sobrenadante se precipitaron con sulfato amónico (65% de saturación) y se resuspendieron en el tampón A (Tris-HCl 50 mM pH 7,5, EDTA 1 mM, 2-mercaptoetanol 7 mM, NaCl 0,15 mM, glicerol 5% v/v) y se cargaron en una columna de intercambio iónico de fosfocelulosa. De esta manera se eluyó más del 95% de la proteína cargada en la columna. Después se lavó la columna con el tampón A, en presencia de concentraciones crecientes de NaCl. Todas las fracciones eluidas fueron ensayadas *in vitro* para la formación de TP-dAMP. La fracción eluida a 0,6 M NaCl presentó actividad y, por tanto, debe contener el complejo TP-DNA polimerasa del fago. Esta fracción se guardó a -20°C en presencia de 50% (v/v) de glicerol.

17. DETECCIÓN DE PROTEÍNAS CODIFICADAS POR FRAGMENTOS CLONADOS DE DNA

17.1. Técnica de “maxicélulas”

Las proteínas codificadas por las ORFs 4 y 5 del fago Cp-1, presentes en diferentes plásmidos, clonados en la cepa CM12 de *E. coli* se detectaron mediante la técnica de “maxicélulas” descrita por Silhavy y cols. (1984) utilizando [α - 35 S]metionina (1.500 Ci/mmol) (Amersham). La electroforesis en gel de poliacrilamida-SDS se llevó a cabo según se describe en el apartado 15.

17.2. Técnica de transcripción-traducción *in vitro*

La expresión *in vitro* de los genes que codifican la holina y la lisozima se realizó mediante el sistema de transcripción-traducción bacteriana libre de células comercializado por Amersham.

18. PURIFICACIÓN DE PROTEÍNAS POR ELECTROELUCIÓN

Después de separar las diferentes proteínas del fago en un gel de poliacrilamida-SDS de porcentaje adecuado, se siguió la técnica de la tinción reversa descrita por Ferreras y cols. (1993). Se recortó la banda correspondiente a la proteína deseada y se dializó frente a EDTA 50 mM para eliminar el Zn^{+2} usado durante el proceso de tinción y así evitar que se obture la membrana durante la electroelución. La electroelución se lleva a cabo en el modelo 422 Electro-Eluter de BioRad.

19. OBTENCIÓN DE ANTICUERPOS ANTI-gp9* DEL VIRIÓN

La obtención de anticuerpos policlonales dirigidos contra la proteína mayoritaria del virión (gp9*) se llevó a cabo mediante la inmunización de conejos con dicha proteína purificada mediante electroelución tal y como se describe en el apartado 18. Se utilizaron tres dosis de 35 μ g de proteína en solución Tris 25 mM, glicina 192 mM y SDS 0.1%, la

primera con adyuvante de Freund completo (1:1; v/v) y, las dos restantes, con adyuvante de Freund incompleto (1:1; v/v), dejando transcurrir 25 días entre ellas. La administración del antígeno se realizó de forma subcutánea. Dos semanas después de la administración de la tercera dosis, se extrajo sangre al conejo para la obtención del suero. La sangre se incubó a 37°C durante 90 min y después se mantuvo a 4°C durante 12 horas para permitir la formación del coágulo. El sobrenadante se centrifugó a $6.000 \times g$ 10 min a 4°C, para eliminar restos celulares y, por último, se inactivaron las proteínas del complemento durante 15 min a 56°C. El suero se centrifugó a $10.000 \times g$ a 4°C y se conservó a -70°C. Se ensayó la especificidad del suero obtenido, así como la posible reacción inespecífica del suero pre-inmune, mediante análisis por "Western-blot" utilizando diferentes diluciones del mismo. Para evitar la detección de bandas inespecíficas, formadas por la posible reacción cruzada del suero con otras proteínas celulares diferentes de la gp9* o gp9, se preincubó el suero a 37°C durante al menos una hora con un extracto obtenido por sonicación de la cepa apropiada transformada con el plásmido sin inserto.

20. TÉCNICA DE "WESTERN-BLOT"

Una vez separadas las proteínas mediante electroforesis en geles de poliacrilamida-SDS, se transfirieron a membranas de nitrocelulosa (BioRad) según describen Sambrook y cols. (1989). La membrana se saturó con leche en polvo desnatada al 5% (p/v) en tampón PBS (fosfato sódico 10 mM pH 7,4; NaCl 140 mM) mediante incubación a 4°C durante 12 horas. Posteriormente, se incubó a temperatura ambiente con agitación suave durante 4 horas en presencia del suero correspondiente para cada caso. Después de tres lavados de 10 min de duración cada uno con PBS que contenía Tween al 0,1% (v/v), la membrana se hizo reaccionar durante 1 hora a temperatura ambiente y suave agitación con anticuerpos de cabra anti-IgGs de conejo, conjugados con peroxidasa (Jackson Immunoresearch). Finalmente, tras un lavado con tampón PBS, las bandas de reacción de los anticuerpos se visualizaron con peróxido de hidrógeno y 4-cloro-1-naftol (Sigma). En algunas ocasiones se utilizó el Kit de detección bioluminiscente de BioRad.

21. DETERMINACIÓN DE LA SECUENCIA N-TERMINAL DE PROTEÍNAS

La secuencia de aminoácidos del extremo N-terminal de las diferentes proteínas se determinó mediante el método de degradación de Edman, según se describe en Spiecher (1994) utilizando un secuenciador automático modelo 477A (Applied Biosystems).

22. DETERMINACIÓN DEL FENOTIPO Lyt DE NEUMOCOCO

El fenotipo Lyt de colonias de neumococo incubadas en medio sólido se determinó como se describe en García y cols., (1985b). En cultivos en medio líquido se siguió el método descrito en Romero y cols. (1993).

23. ENSAYOS DE ACTIVIDAD LÍTICA

Los extractos sonicados de cepas de *E. coli* o *S. pneumoniae* que expresan una actividad hidrolítica sobre la pared de neumococo se ensayaron siguiendo el método clásico para la determinación de la actividad amidásica de neumococo (Höltje y Tomasz, 1976), con algunas modificaciones. La mezcla de reacción contiene 250 μ l (volumen final) de tampón fosfato sódico 20 mM pH 6,9 a los que se añaden 10 μ l de paredes celulares de neumococo marcadas radiactivamente y de 5 a 30 μ l de la dilución adecuada del extracto que se desea ensayar. La incubación se realizó a 37°C a diferentes tiempos y la reacción se paró mediante la adición de 10 μ l de formaldehído 35% y 10 μ l de BSA 4%. Las muestras se centrifugaron en microfuga a 10.000 $\times g$ durante 10 min. A 200 μ l del sobrenadante de cada muestra se les añadió 3,5 ml de líquido de centelleo (Cocktail Ready Safe, Beckman) y se determinó la radiactividad en un contador de centelleo líquido LKB Wallac (modelo 1219 Racbeta). Una unidad de actividad enzimática se define como la cantidad de enzima que cataliza la solubilización de 1 μ g de paredes en 10 min (Höltje y Tomasz, 1976).

24. DETECCIÓN DE HEMOLISINA

Para realizar este ensayo, se obtuvieron protoplastos de los cultivos de las cepas M31(pLSE1), M31(pAMR21) y M31(pAMR22) utilizando un medio con 25% de sacarosa. La detección de hemolisina se realizó como se describe en Seto y Tomasz (1985).

25. REPLICACIÓN DE DNA *IN VITRO*

25.1. Formación *in vitro* del complejo de iniciación TP-dNMP

En el ensayo se usó como molde un oligonucleótido 29-mer que tiene la secuencia de los extremos 3' del genoma del fago. La mezcla de reacción (25 μ l) contiene, Tris-HCl 50 mM pH 7,5, DTT 1 mM, glicerol 4% v/v, BSA 0,1 mg/ml, $(\text{NH}_4)_2\text{SO}_4$ 8 mM, MnCl_2 1 mM, 2,5 μ Ci del indicado $[\alpha^{32}\text{P}]\text{dNTP}$ (0,25 μ M), 200 ng del oligonucleótido que actúa como molde, y 0,5 μ l de extractos solubles de neumococo infectado con Cp-1 obtenidos según 16.2. Después de 10 min a 30°C, se añadió EDTA a 10 mM (concentración final) para parar la reacción y se filtró a través de columnas de Sephadex G50-80 con SDS 0,1% para eliminar los dNTPs no incorporados. Las muestras se analizaron en geles de poliacrilamida-SDS (12%) y autorradiografía. La cuantificación se hizo por densitometría óptica de las autorradiografías.

Los iones amonio actúan en la reacción para favorecer la formación del complejo TP-DNA polimerasa. Para otros sistemas como $\phi 29$ los ensayos se realizaron con proteínas puras a una concentración final de 20 mM $\text{SO}_4(\text{NH}_4)_2$; en el caso de los extractos de Cp-1, la mayor eficiencia se obtuvo a 8 mM (concentración final), aunque la formación del complejo de iniciación tiene lugar en un rango entre 4-15 mM y se inhibe a concentraciones mayores de 20 mM. Esto puede deberse a que al tratarse de extractos y no de proteínas purificadas, el complejo TP-DNA polimerasa está ya formado y no se necesita, por tanto, una alta concentración de iones NH_4^+ .

25.2. Ensayos de elongación truncada

Las reacciones se llevan a cabo esencialmente en las mismas condiciones descritas para el ensayo de iniciación, pero en presencia de diferentes combinaciones de dNTPs/ddNTP, como se indica en cada caso: 10 μ M dATP, dCTP, dGTP y 100 μ M ddTTP para la elongación truncada con ddTTP; 10 μ M dATP, dGTP y 100 μ M ddCTP para la elongación truncada con ddCTP. Para distinguir la posición de la TP unida a los diferentes productos de elongación se utilizaron geles del 12% poliacrilamida-SDS de dimensiones especiales (360 \times 280 \times 0,5 mm). La eficiencia de elongación fue calculada usando el experimento de elongación truncada realizado en presencia de ddTTP, como la relación entre las moléculas elongadas (banda de elongación) y el total de moléculas iniciadas (banda de iniciación más la de elongación). Se considera un factor de corrección teniendo en cuenta la composición específica de cada molécula de DNA.

III. RESULTADOS

1. SECUENCIA DE NUCLEÓTIDOS DEL GENOMA DE Cp-1

1.1. Estrategia de secuenciación

Como ya se ha citado anteriormente, el bacteriófago Cp-1 fue aislado en nuestro laboratorio a partir de muestras de gargantas de niños sanos de Alcala de Henares (Ronda y cols., 1981). Inicialmente se estudiaron sus características morfológicas y propiedades biológicas relevantes (Ronda y cols., 1981, 1983; López y cols., 1981, 1982; García y cols., 1986c), demostrándose posteriormente que su DNA lineal y de doble cadena contenía una proteína covalentemente unida a los extremos 5' (García y cols., 1983c, 1986b). Más adelante, se elaboró un mapa de restricción del genoma, y se comprobó que este DNA tenía una marcada resistencia a las enzimas de restricción como *EcoRI*, *EcoRV*, *SacI*, *SacII*, *BamHI*, *PstI*, *PvuI*, *ScaI*, *SmaI*, *StuI*, *BglII*, *SalI*, etc. (López y cols., 1984). Por otra parte, los 444 nt del extremo izquierdo y los 520 nt del extremo derecho se secuenciaron por el método de Maxam y Gilbert (Escarmís y cols., 1984). El análisis de la secuencia de estos extremos reveló la existencia de una ITR de 236 nt de longitud, y, además, que los 116 nt siguientes presentaban una similitud del 93%, como ya se ha citado en el apartado 2.2.2. de la Introducción. Esta ITR era de mayor longitud que las descritas para otros DNAs con TP. Por otra parte, también se había determinado la secuencia de la región del genoma que codifica la enzima lítica del fago y las regiones flanqueantes, 315 nt que preceden al inicio del gen de la lisozima y 103 nt en el extremo 3' de dicho gen (García y cols., 1988).

Por tanto, como primer objetivo de nuestro trabajo nos planteamos la secuenciación del genoma de Cp-1, siguiendo diferentes estrategias:

a) clonación de diferentes fragmentos de restricción del genoma en los plásmidos pUC18, pBR322 y pBR325. En la Figura 3, se muestra un mapa físico con algunas de las enzimas de restricción empleadas para subclonar fragmentos de DNA del fago y la localización de los plásmidos recombinantes obtenidos. El plásmido pGG4 que aparece en la Tabla 5 se obtuvo a partir de pGG3, digerido con *HindIII* y religado consigo mismo. Los plásmidos pGG61 y pGG66 se obtuvieron haciendo digestiones parciales del fragmento *HindIII* más pequeño con la enzima de restricción *SspI* (ver apartado 9 de Materiales y Métodos). En la Tabla 5 aparece la relación de plásmidos recombinantes obtenidos y la localización del inserto en el genoma del fago.

b) deleciones con nucleasa BAL31 de los fragmentos *Hind*III, *Hinc*II y *Sau*96I-*Sph*I (Figura 3). En la Tabla 5 aparecen todos los plásmidos recombinantes obtenidos y su localización en el genoma del fago.

c) amplificación por medio de la técnica de PCR de un fragmento de 880 pb del extremo derecho del DNA de Cp-1 (Figura 3). El producto amplificado se purificó del gel de agarosa por la técnica de la β -agarasa, y se procedió como se describe en el apartado 8 de Materiales y Métodos.

d) secuenciación directa sobre el DNA de Cp-1 utilizando oligonucleótidos sintéticos como cebadores de la reacción para comprobar distintas zonas del genoma de lectura dudosa así como para atravesar los sitios de restricción.

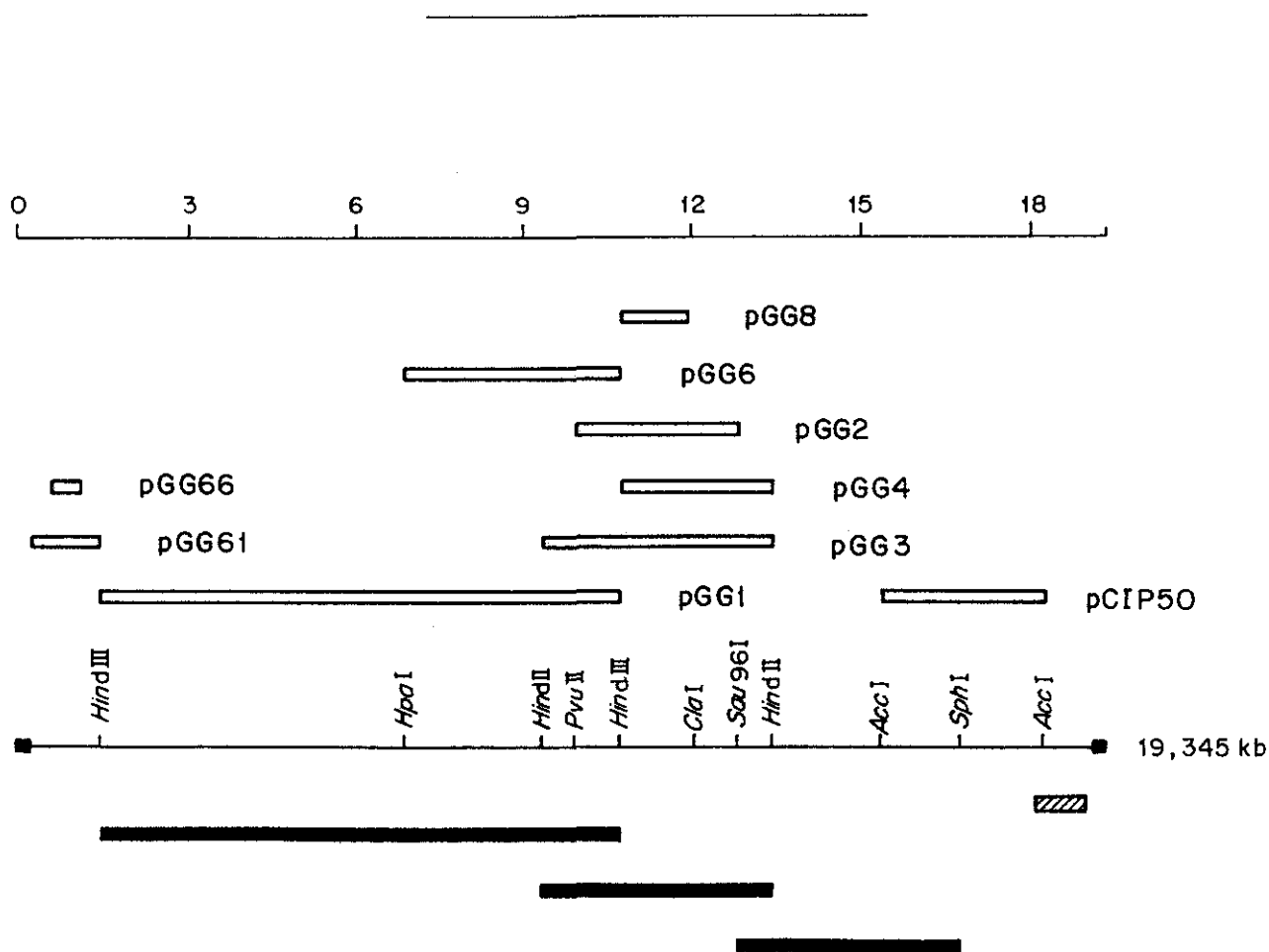


Figura 3. Estrategia de secuenciación del genoma de Cp-1. Las cajas blancas muestran los insertos de los plásmidos recombinantes obtenidos mediante el uso de enzimas de restricción adecuadas. Las cajas negras indican los fragmentos purificados del DNA que se trataron con BAL31. La caja rayada señala el fragmento obtenido por PCR. Las ITRs se encuentran remarcadas en los extremos del genoma.

Tabla 5. Plásmidos recombinantes obtenidos para la secuencia del genoma

Plásmido	Inserto ^a	ORFs ^b	Vector y sitio de clonación
pGG1	1.449-10.736	5-14	pBR322- <i>Hind</i> III
pGG2	9.907-12.821	15-17	pUC18- <i>Sma</i> I
pGG3	9.330-13.438	15-18	pUC18- <i>Sma</i> I
pGG4	10.733-13.438	16-18	pUC18- <i>Hind</i> III/ <i>Sma</i> I
pGG6	6.864-10.736	11-14	pUC18- <i>Hind</i> III/ <i>Sma</i> I
pGG8	10.733-12.066	16	pUC18- <i>Acc</i> I/ <i>Hind</i> III
pGG61	276-1452	1-3	pUC18- <i>Hind</i> III/ <i>Sma</i> I
pGG66	613-1.021	2	pUC18- <i>Sma</i> I
pCIP50 ^c	15.353-18.215	21,22	pBR325- <i>Eco</i> RI/ <i>Hind</i> III
pGG11 ^d	11.437-13.013	17	pUC18- <i>Sma</i> I
pGG12 ^d	9.807-12.874	15-17	pUC18- <i>Sma</i> I
pGG14 ^d	9.961-12.195	15,16	pUC18- <i>Sma</i> I
pGG15 ^d	11.851-12.199	---	pUC18- <i>Sma</i> I
pGG16 ^d	10.229-12.441	16,17	pUC18- <i>Sma</i> I
pGG17 ^d	9.976-12.614	15-17	pUC18- <i>Sma</i> I
pGG18 ^d	9.553-10.964	15	pUC18- <i>Sma</i> I
pGG33 ^d	12.222-12.914	---	pUC18- <i>Sma</i> I
pGG34 ^d	12.214-12.914	---	pUC18- <i>Sma</i> I
pGG9 ^e	5.393-6.554	9	pUC18- <i>Sma</i> I
pGG37 ^e	5.512-8.551	10,11	pUC18- <i>Sma</i> I
pGG38 ^e	6.402-7.905	10	pUC18- <i>Sma</i> I
pGG39 ^e	6.068-7.881	10	pUC18- <i>Sma</i> I
pGG40 ^e	6.628-7.688	---	pUC18- <i>Sma</i> I
pGG41 ^e	5.522-7.545	---	pUC18- <i>Sma</i> I
pGG43 ^e	6.412-8.161	10	pUC18- <i>Sma</i> I
pGG45 ^e	3.623-8.029	6-10	pUC18- <i>Sma</i> I
pAMR45 ^e	3.623-8.029	6-10	pUC19- <i>Sma</i> I
pGG46 ^e	5.543-7.494	---	pUC18- <i>Sma</i> I
pGG52 ^e	5.037-6.016	---	pUC18- <i>Sma</i> I
pGG53 ^e	5.456-6.725	9	pUC18- <i>Sma</i> I
pGG54 ^e	5.173-7.556	9	pUC18- <i>Sma</i> I
pGG55 ^e	5.083-8.215	9-11	pUC18- <i>Sma</i> I
pGG56 ^e	5.145-6.362	---	pUC18- <i>Sma</i> I
pGG57 ^e	6.316-8.070	10	pUC18- <i>Sma</i> I
pGG74 ^f	13.048-14.916	---	pUC18- <i>Sma</i> I
pGG76 ^f	13.291-15.950	---	pUC18- <i>Sma</i> I
pGG77 ^g	18.074-18.954	---	pUC18- <i>Sma</i> I

^a Coordenadas del inserto en el genoma de Cp-1.

^b ORFs completas del genoma que contiene el plásmido.

^c Plásmido obtenido por García y cols. (1987).

^d El inserto del plásmido se obtuvo mediante la delección del fragmento *Hinc*II (9.330-13.438) del genoma con BAL31.

^e El inserto del plásmido se obtuvo mediante la delección del fragmento *Hind*III (1.449-10.736) del genoma con BAL31.

^f El inserto del plásmido se obtuvo mediante la delección del fragmento *Sau*96I-*Sph*I (12.819-16.753) del genoma con BAL31.

^g El inserto del plásmido se obtuvo mediante la amplificación del genoma de Cp-1 por la técnica de PCR.

Los fragmentos de DNA de Cp-1 obtenidos en a) y b) se ligaron al vector digerido con las enzimas de restricción adecuadas (Tabla 5), y con la mezcla de ligación se transformó la cepa TG1 de *E. coli*. Las colonias que contenían plásmidos recombinantes se detectaron por hibridación usando como sonda el genoma de Cp-1. El análisis de los plásmidos para comprobar que llevaban el inserto adecuado se hizo mediante digestiones con enzimas de restricción. Todos los plásmidos obtenidos se purificaron mediante la centrifugación a equilibrio en CsCl-BrEt (Sambrook y cols., 1989) y se comprobaron por hibridación frente a una sonda del genoma de Cp-1, mediante la técnica del "dot-blot". El inserto que contenían se secuenció usando primero los cebadores universales y, después, se avanzó en su secuencia usando oligonucleótidos sintéticos. Estos oligonucleótidos también se usaron, en ocasiones, para secuenciar utilizando como molde el genoma completo de Cp-1.

1.2. Análisis de la secuencia de nucleótidos

La metodología desarrollada en el apartado anterior nos permitió concluir que el DNA lineal de doble cadena de Cp-1 está formado por 19.345 pb (Figura 4). Se han secuenciado ambas cadenas una media de 3,2 veces cada una. El contenido total de G + C es 38,8%, lo que está de acuerdo con el valor previamente estimado para el DNA del fago (41%) y con el del genoma del huésped (39%) (Ronda y cols., 1981).

El análisis de la secuencia del genoma revela la existencia de 29 fases de lectura abierta (ORFs), considerando como tales aquellas que son capaces de codificar más de 50 aminoácidos y que están precedidas por una secuencia de unión al ribosoma (RBS) a la distancia adecuada, y que comienzan por los codones ATG o GTG. De estas ORFs, 23 se encuentran en la cadena superior y las seis restantes en la inferior (Figura 5). El codón de iniciación más utilizado fue ATG y sólo en las ORFs 3, 5, 11 se usó GTG. El codón de terminación más frecuente fue TAA (20 de las 29 ORFs), con TGA en 6 ORFs y TAG en las 3 restantes (Tabla 6). El 93% de los nucleótidos forma parte de alguna ORF, existiendo además siete ORFs solapantes. En los casos de las ORF4-5, ORF12-13, y ORF21-22 el solapamiento es ATGA, donde ATG es el codón de iniciación de una ORF y TGA es el codón de terminación de la ORF que le precede. En el resto de los casos ORF5-6, ORF6-7,

AAAGCATGTACTCCCCACCCCTTTTCAAAAATCCCCCAATGGAAATTTGAAATGAA
 60
 TTTGTCATACAGGGGGTGGGAAAAAGTTTAAAGGGGTAACTTTAACTTTTACTT
 120
 ATGAAGAGAAAAATCCATTTCCACACTAGCCATCTAAAAATCAATCATCTATCATCTT
 TACTTTCTCTTTAGGTAAAGGTGTGATCGGTAAAGATTTTGTAGTAGATAGTAGAA
 CTAAAAATCTTTTATCTGTTCTATTTATCATATATACAAATTTAGAAAAATGCAAGAA
 180
 GATTTTAGGAAAGTAGACAAAGATAATAGTAATATAGTGTAAATCTTTTACAGTTCTT
 ATAAAAATCACTAAATAAATAAAAAATATTTCAAAAATCTCTTACAAATCATTTTAGTAA
 240
 TATTTTAGTGAATTTATTTATTTTATTAAGTTTATAGAGAACTGTAGTAAATCAT
 ATGCTATAATGTTTATAGTTAGAAATGAGATAAAAAATTTCTAAAAATATATAAAAAT
 300
 TACGATATTACCAATATCAATCTTTACTCTATTTTATAAGATTTTATTTAATTTTAA
 AATTTCAAAAAACACTTACAGAGCATTTTGAAGAAATGCTATAATAGTATCAGTTAAAAA
 360
 TTAAAGTTTGTGAACTGTTGTAAGAACTCTTACGATATTATCATAGTCAATTTTGT
 orf1 M K M K A L Y V N H N D E T F
 AAGGAGAAAGAAACATGAAATGAAAGCCTTATATGTAATCACAACGATGAACTTTT
 420
 TTCCTCTTTCTTTGTACTTTTACTTTTCGGAATATACATTTAGTGTGCTACTTTGAAAA
 A E L T H G Q A L D V I F N A D N A E M
 GCAGAACTTACACAGGGCAAGCGCTAGATGTTATCTTAAATGCTGTAAGCTGAAATG
 480
 CGTCTTGAATGTGTGCCCGTTCGOGATCTACAATAGAAATTAACACTATTGOGACTTTAC
 Y G D H V N Q I V D L G N E I K L E L Y
 TATGGGACCATGTAAACAGATTTAGACCTAGGAAATGAGATTAACTTGAATGTAT
 540
 ATACCGCTGCTACATTTGGTCTAACATCTGGATCTTTACTCTAATTTGAACCTAACATA
 R N N E Y I P F A G Y Y E R V K T L S E
 AGAAATATGAATACATTCATTTGCTGTTACTATGAACAGAGTGAACAACTATCTGAA
 600
 TCTTTATTACTTATGTAGGTAAAGCAACATGATAGTCTGCTCACTTTTGTGATAGACTT
 L Y T N I S D W F E E V E * orf2 M
 CTATACACAAATTTTACAGCTGGTTTGAAGAGTTGAGTAAAGGAGATAATAAAATGG
 660
 GATATGTGTTTATAAAGTCTGACCAAACTCTTCACTCATTTCTCTATTATTTTACC
 E L R Q L S V N N I E N L T K K I I D L
 AATTAAGACAATTATCAGTAAACCAATTTGAAATCTTACTAAAAAGATATTGACCTTA
 720
 TTAATTCGTTAATAGTCAATTTGTTGAATCTTTAGAATGATTTTCTAATAACTGGAAT
 K T L T F K P D E F I E S S H V C S E E
 AAAAGCTAAGCTTTAAGCCAGATGAATTTATAGAAAGCTCTATGGTATGTAGCGAGAAC
 780
 TTTGCGATGCAAAATTCGGTCTACTTAAATATCTTTGAGATACCATACATCGCTCCTTG
 L I T S T L K V M Y E F S V K W D L K D
 TTATAACGACCACTTAAAGTTATGTATGAGTTTGTGTTAAGTGGGATTTAAAGATA
 840
 AATATTGGTGTGTGAATTTCAATACATCTCAAAATCACAATTCACCTAAATTTCTAT
 K F Y I D F F Y K I L N N R N D E L H I
 AATTTTACATAGATTTTCTATAAAATTTTAAATATAGAAAGATGAATACACATTA
 900
 TTAATAATGTATCTAAAAAGATATTTTAAATTTATTTATCTTTGCTACTTAAATGTGAAT
 K R L K F I C Y C W L N T E K *
 AACGATTAATAATTTATTTGCTATTGTTGGTTAAATACAGAAATTAATTTACTTTCTTAT
 960
 TTGCTAATTTTAAATAAAGATAACACCAATTTATGCTTTTTATTAATGAAGAATA
 TAAGTTTATCACTTTCCTTTTCTATTTGAAATGTGGTAAACTTAAATGAAGAAATAATAA
 1.020
 ATTCAATAGTGAACGAAAGATAAACTTTTACACCATTTGAATTTATTTCTTTTATTTAT
 TATTTATTTTCTTAAATGTAAGCAAAATTTTAAAGAAAGAGTAAAGCACTTGACAG
 1.080
 ATAAATAAAGAAATTTACATTTTGGTTTAAATTTCTTTCTCATTTTCTGTAAGTGTG
 ATTTACCTATCAATGTAGCTTTAAGCTACGCGAGTAACTTTCAATCTTACAAAAATG
 1.140
 TAAATGGATAGTTTACATCGAAATTCGATGCGTCCATTGAAAGTTAGAAATGTTTTTAC
 orf3 M T I T T E Q V T I N Q R R Q L
 AAGATGGGAATGTACTATCACTACTGAACAGTGACGATAAACCAAGTGCACGCTTC
 1.200
 TTCTACCTTTACACTGATAGTGTACTTTGTTCACTGTATTTGGTTCGACGGTTCGAAG
 P Y E R Y L K E R F K G Y L T I E V V D
 CTTACATTTGAAGCTTATCTGAAGAGCGTTTCAAGGGTATCTCACTATTGAGGTAGTAG
 1.260
 GAATGTAATTCGAATAGACTTTCTCGAAAGTTTCCCAATAGAGTGATACTCCATCATC
 Y E Y K S Y S A S I P E A T A L E Y A E
 ATTATGAATATAAAGTTATTCGGCTCTATCCCTTTGCTACCGCTTAGAATAGCGAG
 1.320
 TAATACTTATATTTTCAATAGCGAAGATAGGGAAGCGATGGCGAATCTTATCGCTC

E Q Q A E G V orf4 M A L T P K Q R K V
 AGGAACAACAAGCGGAAGGGTGAATTAATGGCTTTAACCAACAAACAAAGGAAGTGA
 1.321
 TCCTTGTGTTTGGCTTCCCGACATTAATTTACGAAATTTGTGTTTGTTCCTTCCAT
 R R D Y L T R K K R T L Q Q G A S N A
 AGACGGGACTATTTAACCAAGAAAGAAAGAACTACACAGCAGGGGCGCTCTAATGCT
 1.381
 TCTGCCCTGATAAATTTGGTCTTTCTTTCTTGTGATGTTGTGCTCCGCGGAGATTACGA
 E I K A F M G G R W N F A G M S D K A L
 GAAATAAAGCTTTTATGGGGCTCGGTGGAATTTTGACAGGAATGAGTGATAAGGCGCTA
 1.441
 CTTTATTTTCGAAATACCGCGACCCACTTAAAGCTTCTTACTCACTATTCCGCGAT
 E R A Y D E I K S K G R T Q V F G N H V
 GAACGTGCTATGATGAGATTAAAGCAAGGTGCTACTCAAGTATTGGCAATCATGTT
 1.501
 CTTGACGGGATACTACTCTAATTTTGTTCAGCATGAGTTTATAAACCGTTAGTACAA
 Y T S D Y V K K A K A W Y G D K F S V E
 TACACTAGTACTATGTAATAAAGCTAAGGCTGCTACGGGATAAGTTTCTGTAGAA
 1.561
 ATGTGATCACTGATACATTTTTCGATTTCGATCCATGCGCTTATCAAAAGACATCTT
 K L T Q G F R S S Q R S E L N R F H S V
 AAGCTGACTCAAGGTTTTCGCTCTTCTCAACGCTCAGATTTGAACGCTTTCATTCAGTT
 1.621
 TTCGACTGAGTTCCAAAGCGAGAGAGTTGCGAGTCTCAACTTGGCGAAGTAAGTCAA
 K E V K E Y R S E R D R E A K E R Y I Q
 AAGGAAGTCAAGAAATACCGCTCAGAACGTGACAGAGAACCAAGAACCGCTATATACAA
 1.681
 TTCCTTCAGTTTCTTATGGGAGTCTTGCAGTGTCTCTCGGTTTCTTCCGATATATGTT
 A L E E M H Y N T R D A G N K A Q E K A
 GCGCTGGAAGAAATGCACTATAATACAGAGACCGAGCAACAGGCGCAAGAAAAGCC
 1.741
 CGCGACCTTCTTACGTGATATTATGTTCTCTGCGTTCGTTTTCGCGCTTCTTTTCGG
 F K S M I S R I R R M S A S N F G A F L
 TTTAAGTCTATGATTTTCAAGCATAGGCGAATGAGTGCAGCACTTTTGGGCACTTCTA
 1.801
 AAATTCAGATACTAAGTGGTATTCGCTTACTCACGGTCTTGAACCCCGTAAAGAT
 T G G R S D K V S F D N V M V F V D T D
 ACAGCGGACGCTTGACAAAGTTTCAATTTGATGATGCTGCACTTTTGGGCACTTCTA
 1.861
 TGTCCGCTGCCAGACTGTTCCAAAGTAACTATTACAGTACCAAAACATCTGTGCTG
 G K E T A F E F Q D S L A R E I L D N V
 GGTAAAGAAAGCGCTTTTGAATTTGAGGACAGTGTAGCCGTGAAATCTTGATTAATGTA
 1.921
 CCATTCCTTTGCGAAGAACTTAAAGTCTGTGAGTCCGGCACTTTAGAACTATTACAT
 D K F S K Q F V S D M R R R K K R G K K
 GATAAGTTTTCAGCAATTTGTTTCTGATATGAGAGACGAAGAGCGAGGAGAGAG
 1.981
 CTATTCAAAAGGTTGTTTAAACAAAGACTATACTCTTCTGCTTTCTTCTGCTCCCTTCTC
 orf5 M T C Y Y A G D F E T T T N E E E T E V
 TGACTTGTCTACTATGACGGGACTTTGAAACCACTACAAACGAGGAAGAACAGAGGTAT
 2.041
 ACTGAACGATGATACGTCGCTGAAACTTTGTTGATGTTTGTCTCTTCTTGTCTCCATA
 W L S C F A K V I D Y D K L D T F K V N
 GGCTATCTTGTCTTGAAGTTATGACTATGACAGTTAGACACATTCAGGTAAACA
 2.101
 CCGATAGAAGCAAGATTTCAATAACTGATAGTGTCAATCTGTGTAAGTTCCATTTGT
 T S L E D F L K S L Y L D L D K T Y T E
 CTAGCGTAGAGGATTTCTAAATGCTCTATCTGACCTAGACAAGCTTATACGGAGA
 2.161
 GATCGGATCTCTTAAAGATTTTAGCGAGATAGAACTGGATCTGTTCTGAATATGCTCT
 T G E D E F I I F F H N L K F D G S F L
 CGGGAGAAGATGAATTTATCATATTCTTCAATCTCAAGTTTGACGGCTCTTCTTGT
 2.221
 GCGCTCTTCTACTTAAATAGTATAAGAAAGTATTAGAGTTCAAACTGCGGAGAAAGACA
 L S F F L N N D I E C T Y F I N D M G V
 TATCTTCTTCTGATAATGATATAGAAATGACTTACTTTTAAACGATAGGGGGTAT
 2.281
 ATAGAAGAAAGACTTATTACTATATCTTACATGAATGAATTAATTGCTATACCCCAT
 W Y S I T L E F P D F T L T F R D S L K
 GGTATTCTATTACGCTAGAAATTTCCAGACTTTACGCTAACTTTAGAGATCTTTAAAAA
 2.341
 CCATAAGATAATGCGATCTTAAAGGTGGAATGCGATTGAAATCTCTAAGAAATTTT
 I L N F S I A T M A G L F K M P I A K G
 TCCTTAATTTCTCAATCGCTACGATGGCTGGAGTGTAAATGCCAATGCTAAAGGGA
 2.401
 AGGAATTAAGAGTTAGCGATGCTACGACCTGACAAATTTACGGTTAAGGATTTCCCT
 T T P L L K K K P E V I K F E W I D Y I
 CTACACCTTTGCTAAACATAAGCCAGAGTGAATTAAGCCAGAAATGAGTACTACATTC
 2.461
 GATGTGGAAGAGATTTGTATTGCTCTTCACTAATTCGGTCTTACCTAAGTATGTAAG

H V D V A I L A R G I F A M Y Y E E N F
 21 ATGTAGACGTTGCAATCTAGCCGCTGGTATCTTTGCAATGTACTATGAAGAAATTTTA 2.580
 TACATCTGCAACGTTAGGATCGGGACCATAGAAACGTTACATGATACCTCTTTTAAAAAT
 T K Y T S A S E A L T E F K R I F R K S
 81 CAAAGTACACATCAGCAAGTGAAGCGCTGACCGAGTTTAAACGAGTTTTCAGGAAGTCTA 2.640
 GTTTCATGTGTAGTCTTCACTTCGCGACTGGCTCAAAATTTGCTTAAAGTCTTCAGAT
 K R K F R D F F P I L D E K V D D F C R
 41 AACGAAATTCAGAGACTTTTTCCCAATCTTAGATGAAAGGTGGACGACTTTTGCCGTA 2.700
 TTGCTTTTAAAGTCTCTGAAAAGGGTTAGAACTACTTTTCCACCTGCTGAAAACGCGAT
 K H I V G A G R L P T L K H R G R T L N
 01 AGCATATCGTGGGGCTGGACGTTTGCCCAACCCATAAACACAGGGGGGGAACCTTGAACC 2.760
 TCGTATAGCACCCCGACCTGCAACCGTTGGGATTGTGTCCCGCGCTGGAACTTGG
 Q L I D I Y D I N S M Y P A T M L O N A
 61 AGTTGATAGATATTTATGACATTAACAGTATGTACCTGCTACTATGTCTACAGAACGCTT 2.820
 TCAACTATCTATAAATCTGTAATGTCTACATACGGAAGATGATACGATGCTTGGCA
 L P I G I P K R Y K G K P K E I K E D H
 21 TGCCAAATCGGAATACCGAAACGATACAAAGGCAAGCCCAAGAGATAAAGGAAGACCACT 2.880
 AOGGTTAGCCTTATGGCTTTGCTATGTTTTCGTTCCGTTTCTCTATTTCCTTCTGCTGA
 Y Y I Y H I K A D F D L K R G Y L P T I
 81 ATTATATCTACACATTAAAGCCGATTTTGACTTAAACGTTGGCTACCTCCCAACTATCC 2.940
 TAATATAGATGGTGAATTTTCGCTAAAACGTAATTTTGACCGATGGAGGGTTGATAGG
 Q I K K K L D A L R I G V R T S D Y V T
 41 AGATTAAGAAAAAAGTGGACGCTTTAAGAAATCGGAGTCCGCACTAGCGACTATGTGACTA 3.000
 TCTAATCTTTTGGACCTGCGAAATTTCTAGGCTCAGGCGTATGCTGATACACTGAT
 T S K N E V I D L Y L T N F D L D L F L
 01 CATCGAAGAACGAGGTTATAGATTATATTGACTAATTTTGACCTTGATTATTCTTAA 3.060
 GTAGCTTCTTGCTCCAAATATCTAAATATAAACTGATTAAACTGGAACATAAAGAAT
 K H Y D A T I M Y V E T L E F Q T E S D
 61 AGCACTATGACGCAACTATCATGTATGTTGAAACACTTGAATTTGACAGAAATCGGACT 3.120
 TCGTGATAGTGGTGTAGTATACATAAATTTGGAACCTTAAAGTCTGCTTATGCTGTA
 L F D D Y I T T Y R Y K K E N A Q S P A
 21 TGTTTGATGATTATATCAAACTTACAGATACAGAAAGAAAAACGACAAAGCCAGCAG 3.180
 ACAAACTACTAATATAGTGTGAATGTCTATGTTCTTTTTCGCTGTTTCGGTCTGTC
 E K Q K A K I M L N S L Y G K F G A K I
 81 AAAAAAAGGGCTAAGATTATGCTAAATAGCTTATACGGGAAGTTTGGCGCTAAAAATCA 3.240
 TTTTGTGTTTCGATTCTAATACGATTATCGAATATGCCCTTCAAACCGGATTTTAGT
 I S V K K L A Y L D D K G I L R F K N D
 41 TATCTGTTAAGAACTAGCCTATCTGACGATAAAGGTATATTACGCTTTAAAAATGAGC 3.300
 ATAGACAATCTTTGATCGGATAGACCTGCTATTTCATATATGCGAAATTTTACTGC
 D E E E V Q P V Y A P V A L F V T S I A
 01 ATGAAGAAGTACAGCCGCTTATGACCCGTTGGCTATTGTGAATCTATTGCCC 3.360
 TACTTCTTCTCATGTGCGGCAATACGTTGGGCAACCGGATAAATTTAGATAGAGGG
 R H F I I S N A Q E N Y D N F L Y A D T
 61 GTCACTTTATCATCTCAACGCAAGAACTATGATAATTTCTATATGCTGATACAG 3.420
 CAGTGAATAGTAGAGTTTGGCTGTTCTTTGATACATTAAAGAAATATACGACTATGTC
 D S L H L F H S D S L V L D I D P S E F
 21 ACAGCTTGCACTTGTTCATTCTGACAGCTTGTCTTGACATTGACCGCTCAGAGTTG 3.480
 TGTGCAACGTTGAACAGGTGAAGCTGTGGAACAGGAACGTGAACCTGGGAGTCTCAAGC
 G K W A H E G R A V K A K Y L R S K L Y
 81 GCAAGTGGGCGCATGAGGGGAGAGCGGTTAAGGCAAAATATTACGCTCTAAACTGTATA 3.540
 CGTTCACCGGCTACTCCCTCTCGGCAATTCGCTTTTATAAATGCGAGATTGACATAT
 I E E L I Q E D G T T H L D V K G A G M
 11 TAGAGGAAGTATACAGAAGACGGGACACGACCTAGACGTTAAGGGGCGAGGTATGA 3.600
 ATCTCCTTGACTATGTTCTCTGCGCTGTTGGTGGATCTGCAATTCGCGCTCCACTACT
 T P E I K E K I T F E N F V I G A T F E
 21 CCCCAGAAATCAAGGAAAAATCACTTTTGAACACTTTGTTATCGGGGCAACCTTTGAGG 3.660
 GGGGCTTTAGTTCCTTTTATGTAACACTTTTGAACAAATAGCCCGTTGGAACCTCC
 orf6 M K Q P L K
 61 GTAAGCGTGAAGTGAAGCAGATTAAAGGGGTACGCTGATTATGAACCAACCTTTAAAA 3.720
 CATTGCGACGTTTATTGCTTAATTTCCCCATGCGACTAAATACTTTGTTGGAATTTT

S E K Q T I L F D G F I L S V Y R S F F
 I R E T D Y L V *
 3.721 TCAGAGAAACAGACTATCTTGTGATGGCTTTTACTCTCGGTTACCGCTCTCTTTTT 3.780
 AGTCTCTTGTCTGATAGAACAACTACGAAATATGAGAGCCAAATGGCGAGGAAAAA
 K N L L L H K Q Q V K N K K G Y Y Y Q K S
 3.781 AAAAAATTTATTACATAAACACAGGTTAAAAATAAAAAAGGGTACTACTACAGAAATCA 3.840
 TTTTAAATATGATTGTTGTCCTAATTTTATTTTTTCCCATGATGATGCTTCTTAGT
 N N A P K N T I F L K S Y L K A H Y A Y
 3.841 AACACGCACTAAAAATACTATTTTCTAAAGTCTATCTAAAGCACATTATGCTCTAT 3.900
 TTGTTGGTGGATTTTTATGATAAAAGATTTCAGGATAGATTTTCGTGTAATACGGATA
 E D F D Y I M E L Y K F V S Q E F D K I
 3.901 GAAGATTTTGACTATATCATGGAATTATACAAATTTGTTTCAAGAAATTTGATAAAAT 3.960
 CTCTAAAAGTATATAGTACCTTAATATGTTTAAACAAAGTGTCTTAAACTATTTTAA
 S I N A F Y N L C N Y L E E N K I Y S L
 3.961 TCAATCAATGCTTTTCAATCTTTGCAACTATTAGAGAAAAATAAATCTATTCTCTT 4.020
 AGTTAGTTACGAAAAATGTTAGAAACGTTGATAAATCTTCTTTTATTAGATAAGAGAA
 S S N A L Y D C Y E K S K N R Q N D L E
 4.021 TCTTCTAACGCTCTTTAAGATTGTTACGAGAAATCCAAAAATCGTCAAAACGATTAGAA 4.080
 AGAAGATTGCGAGAAATGCTAACAAATGCTCTTTAGGTTTTAGCAGTTTGTCTAAATCTT
 N L N T I I T P L K F L K S T N G D K Q
 4.081 AATCTCAATACATCATCACCAATTAATTTTAAATCAACAAACGAGATTAACAA 4.140
 TTAGAGTTATGTTAGTAGTGTGTAATTTTAAAAATTTTAGTGTGTTGCTCTATTGTT
 orf7 M A K Q A P K H E N F D T V V A Q A T I
 N G *
 4.141 AATGGCTAAACAGCACCAAAACATGAAACCTTTGATACAGTTGTGGCTCAGGTACAA 4.200
 TTACCGATTGTTGCTGTTTGTACTTTTGAACATATGTCACACGAGTCCGATGTTA
 T A T S N K S D G K Y K Q K K A T K A V
 4.201 CACAGCAAGCTTAACAAATCAGACGCAAGTATAAGCAAAAGAAAGCGACTAAAGCGGT 4.260
 GTGTGTTGACAGATTGTTAGTCTGCGCTCATATTGTTTTCTTGTGCTGATTTCGGCA
 Y L V P A T E E D T K K L V D F G L T L
 4.261 TTACCTTGTCCGACGACAGAAAGACACTAAAAAGCTGTAGACTTTGGCTTGACTTT 4.320
 AATGGAACAGGGTGTGCTCTTCTCTGTGATTTTTGAGCATCTGAAACCGAAGTGA
 Y T P D T E K D P D A K P Y F I V K A T
 4.321 ATACACACGACACGAGAAAGACCCAGACGCAAAACCTTACTTTATTGTTAAGGCAAC 4.380
 TATGTGTGTTCTGTGCTCTTCTGCGTCTGCGTTTGGAAATGAAATAACAAATTCGTTG
 E N V K I F T S E T D F E E V N F G V S
 4.381 CGAAATGTGAAAAATTTTCAAGTGAACGAGCTTTGAAGAGGTCAACTTTGGGGTATC 4.440
 GCTTTTACACTTTTAAAGTGTTCCTTTGCGTGAACCTTCTCAGTTGAACCCCATAG
 Y E D V N P E T G E I T V K K T P N Y K
 4.441 TTATGAAGACGTAAACCCAGACAGGAGAAATCACAGTCAAGAAAGCCGACACTACAA 4.500
 AATACCTCTGCATTGGGCTCTGCTCTTTAGTGTGAGTTCTTCTGGGCTGATGTT
 T E E P V H V A I M F V E G G D N G N D
 4.501 AACAGAAAGCCGTACACGTTGCGATTATGTTTGTGAGGGTGGCGACAAATGGAAACGA 4.560
 TTGCTCTTCTGGGCTGTCACGCTAATACAAACACTCCACGCTGTTACCTTTGCT
 F F R L N A L M M A D T L T L E E V Q P
 4.561 CTTTTCGCTTAAATGCTCTTATGATGGCTGACAGCTTGACCTTGAAGAGTGAACCC 4.620
 GAAAAAGGCAAGTAACAGAACTACCGACTGTGCAACTGGGAACCTTCTTACGTTGG
 V N P F A G L F G K *
 4.621 CGTGAACCGCTTCTGCTGATTGTTTGGAAATAAAAAAGGCTTCCCAATCGGAAAGC 4.680
 GCACCTGGGCAAGCACTAACAAACCTTTTATTTTTCTCGGAAGGGTTAGCCCTTTCG
 TCCAATTATAAGCGTTTTCATGCTTGAAGAGTCAAGTGGTTAGAACTGACGACCA
 4.681 — ↑ 4.740
 AGGTTAATATTTCGCAAAAGTACCGAACTTTTCACTCAACCAATCTTACTGAGCGTGT
 GCAAGCACCCCTTGGGTTGAACCTCTTGGCAGCACTAGACAAGCCTTGAAGAGCCTTA
 4.741 CGTTCGTTGGGAACTCCACATTGGTAGAACGCTGCTGATCTGTTGGAACCTTTTCCGAAT 4.800
 CAGGTTTATCATATCACTTGTCTTTTGTCAAGTATGATATCTTTTGTAAAAAT
 4.801 — ↑ 4.860
 GTCCAAATAGTATAGTATGAACGAAATAAACAGTTTCTACTATATGAAACAAATTTTA
 orf8 M T S Q E C L A V L D A A M A
 4.861 TGAAGAGAGAGGCTATGACCTCAGGAATGCTCGAGCTGTAGAGCTGCAATGGCA 4.920
 ACTTTCTCTCCGATAGTGGAGTGTCTTACGGAGGTCACGATCTGCGAGTTACCGT

K V G N D E E I E S L T A D L I D I K A
AAAGTCGGAAACGATGAAGAAATTGAGAGCCTAACCGCTGACTTGATTGACATTAAAGCT
921 TTTCAGCCTTTGCTACTTCTTTAACTCTCGGATTGCGACTGAACTAACTGTAATTCGGA 5.080
F V G E I D T V V S V L N E D V E R P N
TTTGTTCGGAGAAATTGACACAGTTGTCCTCAGTCTTGAATGAAGACGTTGAGCGGCCAAAT
981 AAACAGCCTCTTTAACTGTGTCAACAGAGTCAGAACTTACTTCTGCAACTCGCGGGTTTA 5.040
L K N G N L R S A N N E L Y R R L G Q Q
CTTAAAAACGGTAACCTTACGTTTCGGCTAATAACGAACTTTACGCGCGTTTAGGGCAACAA
041 GAATTTTGGCATTGAATGCAAGCCGATTATTGCTTGAATGGCGGCAAAATCCCGTTGTT 5.100
D E I M K Q A Q E D M S V V S A I N A V
GATGAAATCATGAAACAAGCACAGAAGACATGAGCGTAGTATCAGCAATCAATGCTGTT
101 CTACTTTAGTACTTTGTTGCTGTTCTTCTGCTACTCGCATCATAGTCGTTAGTTACGACAA 5.160
I *
ATTTAAAGAAAGGAAGAAAGAAAGATGAAACCATTTCCTAAAAAGCATTAACTGTTATCCT
161 TAAATTTCTTCTCTTCTTTCTTCTACTTTGGTAAAAAGTTTTCGTAAATTGACCATAGGA 5.220
AACAACGCACTAGACGCACTCAAGGACGAAACAGTGGCAGAAGTTACTTCGCCA
221 TTGTTGCGTGATCTGCGTGAGTTCCTGCTTGGTCTTGTGACAGGCTCTCAATGAGGCGGT 5.280
GCAGTTATGCGTGCTGACACACCGACGCGAGGAAGTGCCAACTACCCAGCGCAAGCCAG
281 CGTCAATACGAGCACTGTGTGGTTCGCGTCTTACGCGTTTGATGGGTGCGGTTGCGGTC 5.340
CCAGCGAAGTTAGGGCGGTAGAAATGAACATCGACCACGAAACATTGTGAAGAGGGAG
341 GGTGCTTCAACTCCGCGATCTTTACTTTGATGCTGGTGCTTTTGAACATCTTCTCCCTC 5.400
oxf9 M A N K I T T F L S G Q T G K Q I
AAGAATAGAATGCTAATAAAATTACCCTTTCTTATCAGGACAACTGGTAAACAAAT
401 TTCTTATCTTACGATTATTTTAAATGGTGAAAGAATAGTCCTGTTGACCATTTGTTTAA 5.460
S N I D L L N S I R T R A S A D Y Q A D
TCAACATTGACCTATTGAACCTCTATCCGACCGGTGCAAGTGCTGACTATCAGGCAGAT
461 AGTTTGTAACCTGGATAACTTGAGATAGGCGTGGGCAAGTTCAGCACTGATAGTCGCTCTA 5.520
I P V L E G A R I N H A T V P Y Q D F Q
ATTCCAGTTCTTGAGGGCGCAGCATTAAACCAAGCAACAGTTCATATCAGGATTTTCAA
5580 TAAGGTCAAGAAGTCCGCGGTGCGTAATTTGGTGCGTTGTCAAGGTATAGTCTCTAAAGTT
K H A N E F F T A L V N R I G S T V I K
AAGCAGCGTAAATGAGTTTTTACAGCCTTGGTCAATGATGATGGGTCAACGGTTATCAAG
581 TTGTTGCGATTACTCAAAAAGTGTGCGAACAGTTAGCATAGCCCACTTGGCAATAGTTC 5.640
A L T Y E N P L A I F K S E T F E F G D
GCGCTTACTTATGAAATCCGCTTGCTATTTTCAAGTCTGAAACCTTTGAGTTTGGGGAC
641 CGCGAATGAATACTTTTAGGGGAAAGATAAAGTTTCAGACTTTGGAAGTCAAAACCCCTG 5.700
T L Q E I Y V H P A E K K T Y D A K S D
ACATTGCAAGAAATCTATGTACACCCAGCGGAAAGAAACCTATGACGCTAAGTCAGAC
701 TGTAACGTTCTTTAGATACATGTGGGTGCGCTTTCTTTTGGAATGCGGATTGAGTCTG 5.760
V S P E K F A D T D I E A F Y H T L N N
GTAAGCCCGTTTAAATTCGCTGATACAGACATCGAAGCCTTTTATCACACTTTGAACAT
761 CATTCGGGCAAAATTTAAGCGACTATGTCTGTAGCTTCGGAAAAATAGTGTGAAACTTGTTA 5.820
E N Y Y E R T F E R A W I Q K A F V S D
GAAACTACTATGAGCGTACCTTTGAACGTGCTTGGATTGAGAAAGCCTTTGTTTCTGAC
821 CTTTGTGATGATCTGCGATGGAACCTTGCAAGAACCTTAAGTCTTTGGAACAAAGACTG 5.880
A G F D E F V G D K M F T S S L S S D T L
ATGGCTTTGACGAGTTTGTGGACAAATGTTTACATCATTGCTATCATCTGACACCGTG
881 TACGGAAGTCTGCTCAACACCTGTTTACAAATGTAGTAACGATAGTAGACTGTGCGAC 5.940
D E Y Q A V R V Y L R N H L R K S L I Q
GAGAGTACCAAGCAGTTAGGGTGACTTGAGAAATCACTTGGGAAGTCTCTTATACAG
941 CTGCTCATGGTTGCTCAATCCACATGAACCTTTTGTAGTGAAGCGCTTCAGAGAAATATGTC 6.000
* K R F D R I C
T L K G N D K K I T V A G T K I D E T K
ACCTTGAAAGGTAAACGATAGAAATTAACGTCGCGAGTACGAAGATTGACGAAACAAAA
001 TGGAACCTTCCATGCTATTTCTTTTAAATGGCACCGTCCATGCTTCTAACTGCTTTGTTTT 6.060
V K F P L S L F I V T A P V F I S S V F
Q D F V V D F N Q S L I N L S K R F I
CAGGACTTTGTCGTAGACTTTAACCGTCTCTAATTAAGTGTCTAAAGTTTACAAAT
061 GTCTGAAACAGCATCTGAAATTTGGTCAGAGATTAATGAACAGATTGCAAGATGTTAA 6.120

P S R T T T T F N M P V G V P N M N T A I E D
 CCAAGCGGTACAAAGCTTTAAACAACCGCTGAGGCGTGCCAAACATGACGGCGGATGGAAGAC
 6.121 GGTTCGGCATGTTGCAAAATGTTGGGACATCGGCACGGTTTGTACTGCCGCTAACTTCTG 6.180
 G L R V V N L L G T P T G F M V A I S S
 Q Y L V I S A E F S T H L D M L L A N A
 CAATATCTGGTTATCTCAGCAGAATTTCTACACATCTTGACATGTACTGTGCTAAACGCC
 6.181 GTTATAGACCAATAGAGTGTCTTAAAGATGTGTAGAAGTGTACAATGAACGATTGCGG 6.240
 W Y R T I E A S N E V C R S M N S A L A
 F N M D K A S V L A R T I V V D D F E K
 TTTAACATGGATAAAGCAAGCGTACTCGCTCGCACAAATGTAAGTAGATGATTTTGAAAAA
 6.241 AAATGTACCTATTTGTTTCGATGAGGAGCGGTGTAGCATCATCTACTAAACATTTTT 6.300
 K L M S L A L T S A R V I T T S S K S F
 F T G E G A N N G R K P V A F L I S A K
 TTCAGGGGAGGGCGCAACAATGGACGCAAGCCAGTTGCTTTCCTTATTTCTGCTAAG
 6.301 AAGTGCCTCTCCCGGTTTGTACCTGCGTTTCGCTCAACGAAGGAATAAAGAAGATT 6.360
 N V P S P A F L P R L G T A K R I E A L
 S I I N K D K L V H M E A I R N P R N M
 TCTATCATTAAACAAGATAAACTGGTACACATGGAAGCAATCGCAACCCCTCGCAACATG
 6.361 AGATAGTAATTTGTTTCTATTGACCATGTGTACCTTCGTTAGGCGTTGGGAGCGTTGTAC 6.420
 L S L T K I N W Y T W K Q S A T L A T M 4-orff
 T Y N Y F Y H H Y M T S L F E N I
 ACCTCAACTATTCTACCATCACCACATGACTGACCGCTTCACTTTTGAAAAATT
 6.421 TGGATGTTGATAAAGATGGTAGTGGTGTACTGGTGGGAAAGTAAAAACTTTTGTA 6.480
 H F W Y V E E A *
 CATTTCTGGTATGTTGAGGAAGCCTAAAGGCTGACCAAGGGGGCAATAGCCCGCTAT
 6.481 GTAAAGACCATACAACTCCTTCGGATTTCGACTGGTTCGCGCGTTATCGGGCGGATA 6.540
 orf10
 TTTATTAAGTGAAGGGGACTAAATGAGTTACAAGAAATACAAGGACATCTGGTAAGA
 6.541 AAATAATTCACTTTCCCGTGATTTACTCAATGTTCTTAATGTTTCGCTGTAGAACCATTCT 6.600
 I E L N K E T V E R N R L A F F E F Y F
 TTGAGCTAAACAAGAAACCGTAGAGGTAACCGCTAGCGCTTTTGTAGTTTATTTCGA
 6.601 AACTCGATTGTTTCTTTGGCATCTCGCATTTGGCAGATCGGAAAAAATCAAAATAAAGT 6.660
 N Y F Y N I V V N Y F T W E G L P N D I
 ATTATTTCTATAATATGTTGTAACCTATTCTACTTGGGAGGTTTGGCTAATGACATTG
 6.661 TAATAAGATATTATAGCACCAATTGATAAAGTGAACCTCCCAACCGATTACTGTAAC 6.720
 D E L F I E K K L I E N G H V A F F H D
 ACGAGTTGTTTATAGAAAAGAGCTGATAGAAAAATGGGCATGTAGCTTTTTCATGATG
 6.721 TGCTCAACAAATATCTTTTCTCGACTATCTTTTACCGTACATCGAAAAAAGTACTAC 6.780
 D T F G Y I A Q G G T R G E R L N H Y D
 ACACTTTGGCTATATCGCACAGGTGGAACACGAGGGGAACGCTTAAACCAATTATGACC
 6.781 TGTGAAAACCGATATAGGTGTTCCACCTTGTGCTCCCTTGGGAATTTGGTAATACTGG 6.840
 Q P L T Y Q P V N A S S M N Y F K Q M E
 AGCCTTTGACCTATCAACCGGTAAACGCTAGCAGTATGAACATTTTAAACAAATGGAAA
 6.841 TCGGAAACTGGATAGTTGGGCAATTCGATCGTCATACTGTATAAAATTTGTTTACCTTT 6.900
 I A Y T E N D F R V I E E L H K D N P D
 TCGCTTATAGTAAATGATTTTAGAGTTATTGAAGAATTACACAAGGATAATCCAGACA
 6.901 AGCGAATATGACTTTTACTAAAATCTCAATAACTTCTTAATGTGTTCCTATTAGGTCTGT 6.960
 K I K R P C I V I P N N N F Y E P Y I G
 AGATTAAGAGACCTGTGTTTGTGATTCTCAACAATAATTTCTATGAGCGGTATATAGGCT
 6.961 TCTAATTTTCTGGAACATAACACTAAGGATTGTTATTAAAGATACTCGGCATATATCCGA 7.020
 Y L E L F C E K L A D I E L T I Q L N R
 ATTTAGAGTTATTTTGGGAAAAGTTGGCAGATATTGAGCTTACGATTCAATTAATAGAA
 7.021 TAAATCTCAATAAAACGCTTTTCAACCGTCTATAACTCGAATGCTAAGTTAATTTATCTT 7.080
 N A Q I T P Y F I F A D N T N V L S M K
 ATGCACAAATTACACGCTATTTTCATCTTTGCGGATAATACAAACGCTGCTATCTATGAAAA
 7.081 TACGTGTTTAAATGCGCATAAAGTAGAAACGCCATTATGTTTGCAGATAGATACTTTT 7.140
 N I F N K I A N F E P V G Y L N K Q K D
 ACATCTTTAATAAGATTGCAAAATTTGAACCGGTGTGTATCTCAACAGCAGAAAGACC
 7.141 TGTAGAAATATTCTAACGGTTAAAACCTTGGGCAACACATAGAGTTGTTGCTCTTCTG 7.200
 Q D G Q D S F K Q L S D Y I Q V F R T D
 AAGACGGACAAGACAGCTTTAAACAGTTGTGACACTATATCCAAGTATTCCGAAACGGACG
 7.201 TTCTGCGCTGTTCTGTGCAAAATTTGTCAACAGTCTGATATAGGTTTATAAGGCTTGCCTGC 7.260

61	A P F L L D K L H D E K L R V M N Q L L CACCGTTTTACTTGACAAGTTGCAGATGAGAAATACGAGTGTGACCGATTGCTGA	7.320	8.461	Q S I K A L L F I R F K I V G A Y Y S F CAATCCATTAAAGCGTTGTTTATTCATTOGCTTTAAGATTGTGGGAGCTACTATAGCTTT	8.520
	GTGGCAAAATGAACCTGTTCAACGTGCTACTCTTTAATGCTCACTACTTGGTCAACGACT			GTTAGGTAAATCGGCAACAATAAGTAAGCGAAATCTAACACCCCTCGCATGATATCGAAA	
21	T F I G I N N N P S D K K E R L V V S E CTTTCATTGGGATAAATAACACCCATCAGATAAAAAAGAGCGTCTGGTAGTCTCGAAG	7.380	8.521	E K L T S K S K L K G F G R V I D D N N GAAAACTGACCGAGCAAGCTGAAGGCTTTGGCAGAGTATAGACGATAACAAAC	8.580
	GAAAGTAACCTATTATTGTTGGGTAGTCTATTTTTCTCGAGACCATCAGAGGCTTC			CTTTTTGACTGGTGGTTTTGGTTCGACTTTCCGAAACCGTCTCACTATCTGCTATTGTTG	
81	A I S N N G V I S A N I E V G W K S R R CTATTTCTAATAATGGGGTTATCTCAGCGAATATTAGGTAGGCTGGAGTCCAGAGAA	7.440	8.581	Y F S R I P L V N E V V H W D N G V I V TATTTTTCAGCATACCCCTTGTAATGAGGTGGTACATGGGATAACGGGGTTATCGTA	8.640
	GATAAAGATTATTACCCCAATAGAGTGGCTTATAACTCCATCCGACCTTCAGGTCTTCTT			ATAAAAAGTGCGTATGGGGAACATTACTCCACCATGTAAACCTATTGCCCAATAGCAT	
41	K E V E L I N K C Y G L E I S V K P A E AATTCGTTGAGCTTATCAATAAATGTTACGGGTGAGAAATCAGGTAACCAAGCGGAA	7.500	8.641	T P N Y Q M N I T G L K E T R V E V D G ACACCTAACTATCAAAATGAACATCACAGGACTAAAAGAAACCGGTGTAGAGGTTAGCGGT	8.700
	TTAAGCAACTCGAATAGTTATTACAATGCCGATCTTTAGTGCATTTTGGTGGCTTTT			TGTGGATTGATAGTTTACTTGTAGTGTCTGATTTTCTTTGGGCACATCTCCAACGCCA	
01	T I Q Q F N L D K V A L D L A E K E G T CGATTGAGCAGTTTAACTTGCAGAAAGTGGCGCTAGACTTGGCAGAAAGGAGGGGACAA	7.560	8.701	Q Q L F E D W S T F K I N V T N D R K G CAGCAACTCTTTGAAGATTGGTCAACCTTTAAATCAATGTAAACCAAGATAGAAAGGGA	8.760
	GCTAAGTCGTAATTTGGAACGTTTTCACCGGATCTGAACGCTTTTCTCTCCCTGTT			GTGCTTGAGAACTTCTAACCGATTGGAAATTTTAGTTACATTGGTGTCTTCTTCCCT	
61	I I D P E * TCATTGACCCAGATAACACAACAGCAAGATTGCAACCTTTTAAATCCAGATATAGA	7.620		orf13 M I I I N V P R T V M T A E R G H E N L *	
	AGTAAGTGGGTCTTATTGTTGTGCTGTCTAAGTTGGAAAAATTTTAGGTCTATATCT		8.761	GTACCTCGCAGTTTATGACAGCAGAAAGGACATGAAACCTATGATAATTATCAACT	8.820
121	orf11 M T G R L D G L A V D E N G E F L H AATCCCGTGACGGGAGCGGTGGAGCGCTTGGCAGTAGATGAAACGGGAGTTTCTGCAT	7.680		CATGGAGCGGTGCAATAGTGTCTTCTCTGCTACTTTTGGATCTATTAATAGTTGA	
	TTAGGGCACTGCCCTGCCGACCTGCCAACCGTCACTACTTTTGGCGCTCAAGAGCGTA		8.821	L S E T T D T L Q I E V T G H G D D T D TGTCGGAGACACGGGACGCTACAAATCGAAGTCACGGGACAGGAGAGGATACAGACC	8.880
181	Y N T I I D Q T Y N E L F K D M E L V N TACAACAGATTATAGACAGACCTATAACGAGTTATTAAAGGACATGGAGCTAGTCAAT	7.740		ACAGGCTCTGTGCTGTGCGATGTTTAGCTTCAGTGCCCTGTGCTCTGCTATGCTCG	
	ATGTTGTGCTAATATCTGCTGAGATATGCTCAATAAATCCTGTACCTCGATCAGTTA		8.881	Q S C A R V S T V C D C I Y L F L K S N AGTCTTGGCCCGGTGTATCAACCGTTTGGGATGTATTACTTATCTTAAATCAACA	8.940
741	G V S D N F K K E F C K H F Y M R E I G GGAGTTTCAGACAATTTCAAGAAAGAGTTTTCGAAACACTTTTACACAGGAGGATAGGA	7.800		TCAGAACGCGGACATAGTTGGCAACGCTAACATAAATGAATAAGAAATTTAGATTGT	
	CCTCAAAAGTCTGTTAAAGTTCTTCTCAAAACGTTTGTGAAATGTTGCTCTCTATCCT		8.941	I D D Y V K K D G Y T L L R I F K K R T TAGATTATGTTTAAAAAGATGTTATACATTGCTACGGATTTTTAAAAACGTAATA	9.000
301	L E T F A R F Q I A L E E V L N N E C F CTTGAGACTTTTCGACGTTTCCAGATAGCCCTTGAGGAAGTTCTAANCAAGAGTGTTC	7.860		ATCTACTAATAACAATTTTTCTACCAATATGTAACGATGCCTAAAAATTTTTGCTATGAT	
	GAACCTGGAAGCGTGCAAGAGTCTATCGGGAACCTCTCAAGATTGTTGCTCACAAG		9.001	T V Q T L K A L F * CGGTACAACTTTAAAAAGCATTATTTTAAAGTTACATCGTAACGCTTGAGCACTTTATAA	9.060
361	N L F K Y L A E I R N K A I K D L N Q S AATCTGTTTAAATATCTTGCTGAAATCAGAAACAGGCTATCAAGGACTTAAATCAGTCT	7.920		GCCATGTTGAAATTTTGTAAATAAATCAATGTAGCATTTGGCAACTCGTTGAAATATT	
	TTAGACAAATTTATAGAACGACTTTAGTCTTTGTTCCGATAGTTOCTGAATTIAGTCAGA		9.061	orf14 M A K T T AAATTCATTAAGGTTTATAATAAAGAAAGAGGTAGAAACAAATGGCAAGACAATA	9.120
921	M N I D T V G N Q K A D G Q A L Q I A N ATGAATATTGACAGGTTTGAACACGAGAAAGCAGACGGGCAAGCCTTCAATGCAAAAC	7.980		TTTAAGTTAAATCCAATATTATTCTTTTCTCCATCTTTGTTTACGGTTCTGTTGAT	
	TACTTATAACTGTGCCAACCTTTGGTCTTTCGTCTGCGCGTTGGAAAGTTTAGCGTTG		9.121	K L V R G I H S W I K F Q K H Q G V E S AACTGTGACGAGTATTCACTCATGATTAAGTTCCAAAAACCAAGGAGTGGAGAGCC	9.180
981	T T P Q E R K E I V F T E R Y G V I E Y ACCACACCCCAAGAGCGGAAAGAAATGCTCTTACTGAGCGCTACGCGGTGATAGAGTAC	8.040		TTGAACATGCTCCATAAGTGAGTACCTAATTCAAGGTTTGTGTTGCTCACTCTCGG	
	TGGTGTGGGGTTCTCGCTTTCTTTAGCAGAAATGACTCGCATGCCGCACTATCTCATG		9.181	L S I E G K Q A L A D L S Q D K N G D T TTCTATTGAGGTAAGCAGGCGCTTGTGCTGCTGACAGGATAAAAACGGGACACAA	9.240
041	A D N L V E N H Q K N N A D T K S N V S GCTGACAACCTTGGTAGAAAACACGAGAAACACGCTGACACAAAAGCAAGCTCTCA	8.100		AAAGATAACTCCCATTCGTGCGGAACGACTGAACAGTGTCTATTTTGCCCTGTGTT	
	CGACTGTTGAACCATCTTTTGGTGGTCTTTTGTGCGACTGTGTTTTTGTGTCAGAGT		9.241	T L I L N A D V D K L N T I H S A I P F CCCTTATTTTAAATGACAGAGTGTGATAAATTAATTAATTAATTAATTAATTAATTAAT	9.300
101	G W S G S S L A E R L Q R N A E L K D I GGGTGGAGCGTTCAAGCCTTGGCAAGCCTTACAACGTAACGCAAGTAAGAGATATA	8.160		GGGAATAAAATTTAAGTCTGCAACTATTAAATTTATGATAAGTAAGTGTATGGAAT	
	CCCACCTGCCAAGTTGCGAAACGGCTTGGCAATGTGCAATTCGCTCTTAATTTCTATAT		9.301	I S V S E E F S G V D P N Q M K T A T I TTTCTGTTTCAGAGAAATTTTCAAGGAGTTGACCCCTAATCAAAATGAAACAGCAACCATCA	9.360
161	Q F Q I F N I C D K L F L Q V F * CAGTTTTCAGATATTTAATATATGTAAGCTGTTTTTACAAGTCTTTTGAAGGTTGA	8.220		AAAGACAAGTCTTCTTAAAGTCTCAACTGGGATTAGTTTACTTTTGTGTTGTTAGT	
	GTCAAGTCTATAAATATATACACTATTGCAACAAAATGTTGAGAAATCTTTCCACAT		9.361	S Q D L T K F P L N N G E L V K F N K E GCCAAGATTAAACAAATTCCTCTAATAATGGGAACTGGTTAAATTTAAACAAAGAA	9.420
221	orf12 M K D L S S A K I L K Y D S M L E E I TAGATGAAGATTATCAAGCGCTAAAATACTAAAATATGATAGTATGTTAGAAAGATC	8.280		CGGTCTTAAATGTTTTTAAAGGAGATTATTAACCCCTTGACCAATTTAAATGTTTCTTT	
	ATCTACTTTCTAAATAGTTCGCGATTTTATGATTTTATACTACTACATACTCTCTTAG		9.421	K D G I S V N D E E L K A D I S Q L I E AAGATGTTATTTCTGTAATGATGAAGAAATGAAAGCAGATATTTCTCACTTATTGAG	9.480
281	T L F S F Q D F A Y S D D G L Y Y I Q S ACGCTTTTCAGCTTCAAGACTTTGCTTATAGTGATGATGGATTGATTATATCCAGTCA	8.340		TTCTACCATAAAGACATTACTACTTCTTAACCTTCTGCTATTAAGAGTTGAATAACTTC	
	TGCGAAAGTTCGAAAGTCTGAAACGAATATCACTACTACCTAACATAATATAGTTCAGT		9.481	A K L T G F P V V L N Y E H I L L D N F CTAAATTTGACAGGTTTCCAGTTGATTAAATTAATGACATATTTTACTGATATATTTT	9.540
341	N S R R L G D L S K L W I K L K P I F Y AATAGCAGAGCGTTAGGCGACTTGTCTAAATTTGGATGATAAATTAAGCCTATCTTTTAT	8.400		GATTTAAGTGTCCCAAGGTCACATAATTTAATCTTGTAATAAATGAACATTTAAAAA	
	TTATCGTCTGCCAATCGCTGAACAGATTAAACACCTATTTAATTTGGATAGAAAAATA		9.541	S K I E T S I N I S T Y Q P F N D F V E CAAAAATGAAACATCTATCAATATTTCCACTTATCAACCGTTAATGATTTTGTGGAAG	9.600
401	H Y E S I E N E T F W T I R K S Y Q P L CATTTAGAAAGCATTGAAATGAAACTTTCTGAGTATCCGCAAGAGCTACCGCGGTTA	8.460		GTTTTTAACTTTGTAGATAGTTATAAAGTGAATAGTTGGCAAAATTAATAAACCTTTC	
	GTAATACTTTCGTAACCTTTACTTTGAAAGACCTGATAGCGGTTTTCGATGGTGGCAAT		9.601	V T L T N P A T N E Q D T V I I S T N E TAACCTTGACCAATCCAGCTACAAATGAACAGACCTGTTATTATTCAACAAATGAGT	9.660
				ATTGAAACTGGTTAGTGTGATGTTTACTTGTCTGTGACAATAAAGTTGTTTACTCA	

L Y N G K M I F N R F V C S V L V E K N
 TATACAAATGAAAAATGATTTTAAATAGATTGTTTGCAGTGTTTAGTTGAAAAAATG 9.720
 561 ATATGTTACCTTTTACTAAAAATATCTAAACAAACGTCACAAAATCACTTTTITAC
 D K H I I D I T V N S H K E A D R I H V
 ATAAACATATTATTGATATACTGTAATTCATATAAGAAAGCTGACAGAAATCCATGTTA 9.780
 721 TATTGTATAATACTATATTGACATTTAAGAGTATTTCTTGAGCTGCTTAGGTACAAAT
 T I K W F T Q Y R G Q K I L D V N E I G
 CCATAAAATGGTTTACACAATACAGAGGACAAAAATCCTTGACGTAATGAAATGGTC 9.840
 781 GGTATTTTACCAAAATGTGTATGTCTCTGTTTTTTAGGAAGTCATTTACTTTAACCAG
 Q D A H F E G Y E N P N V E L T Y S P I
 AAGACGCACACTTTGAGGATATGAAACCTTAACGTAGAATTGACTTATAGTCTTATTC 9.900
 341 TTCTGCGTGTGAAACTCCCTATACTTTTGGGATTGCACTTAACTGAATATCAGGATAAG
 P P A V S T P T E T P S P T P E T T G N
 CACCAGCTGTAAAGCACCAACGAAACACCAAGCCCAACCCCTGAAACAACTGGAACA 9.960
 302 GTGGTCGACATTGCTGTGGTGGCTTTGTGTTGGGTTGGGACTTTGTGACCTTTGT
 T P Q P I A E E S H * orf15 M N
 CACCGCAACCAATGCGAGAAGATCTCACTAAAAAAGAGGATTTTAAACATGAATC 10.020
 961 GTGGCGTTGGTTAAAGCTCTCTAGAGTATTTTATCTTTCTCTAAAAATTTGTACTTAG
 P E E F K D E F F R A Y R G R Y S S Y W
 CAGAAGAAATTAAGATGAATTTTTCAGGCTTACCGTGGAGCTTATCTGCTTACTGGG 10.080
 021 GTCTCTTAAATTTTACTTAAAAAGTCCGAAATGGCACTGCGATAAGCAGAATGACCC
 V E R W G L I P S I P T S F D N A N S I
 TAGAAGCTTGGGGCTTATCCCTCAATCCCTACAGCTTTGATAAGCCCAATTTCTATCT 10.140
 081 ATCTTGCAACCCCGAATAGGGAGTTAGGGATGGTOSAACTATTGCGGTTAAGATAGA
 Y E L L A W L Q R A F K Q L L D D F V A
 ACGAGCTTTTGGCTGGCTACAGGTCCTTTAAGCAGTTGCTTGAAGCACTTTGTGGGT 10.200
 141 TGCTCGAAAAACCGCAGATGTGCGACGGAATTCGTCAGGAACCTGCTGAACACCGCA
 L E S E L E D Y K N A L T E L L E Q L I
 TGGAAAGCGAACTGGAAGACTATAAGAAGCTTTAACTGAACCTCTAGAGCACTTATCC 10.260
 201 ACCTTTGCTTGACCTTCTGATATCTTGGGAAATGACTTGAGGATCTGTTGAATAGG
 P L L I R R Y M E S K E A D D W F N K K
 CTTTGCTTATCGTAGATACATGGAGAGCAAGGAAGCGGACGCTGGTTTAAATAAAG 10.320
 261 GAAACGAATAGGCATCTATGACCTCTGCTTCTTGGCTGCTGACCAAAATATTTTTTC
 A A I Y Y N K I I K P Y I D A E I A K V
 CGGCCATCTACTATAACAGATTTAAGGATTTAAGCGCAGAAATAGCTAAAGTCA 10.380
 321 GCGGTTAGATGATATTGTTCTAATAGTTGCGAATATAACTGCGTCTTTATGATTTCACT
 N K K I A D L E K K V D D E V K R L D G
 ATAAGAAATGCTGACCTTGAAAGAAAGTAGATGATGAAGTTAAGGACTAGATGGAC 10.440
 381 TATTCTTTTAGGACTGGAACCTTTTCTTCTACTACTTCAATTCGCTGATCTACCTG
 R I D A L N D K L E K E I K K L D D R I
 GGATTGACGCTTTAAACAGTAAGCTAGAAAAGAAATCAAGAACTTGACGACAGATCA 10.500
 441 CCTAACTGCGAAATTTGCTATTGATCTTTTCTTTAGTTCTTTGAAGTGTCTTGTAGT
 T Q E V A K L N E R I T A E N N A L K E
 CACAGGAGTCTGTAAGTTAAAGAGCGGTATAACAGCAGAAACACGCACTCAAGGAGA 10.560
 501 GTGTCCTTCAGCGATTCAATTTGCTGCGATATTGCTGCTTTTGTGCGTGAGTTCTCT
 * P S
 R I E A L E N L T Q A C K M L C V K S L
 GAATCGAAGCGCTAGAAAATCTAACGAGGCTTGCAAAATGCTTTGCGTAAATCATGA 10.620
 561 CTTAGCTTCGCGATCTTTTAGATTGCGTCCGAAGCTTTTACGAAAGCATTTTAGTAAC
 F R L A L F D L A F K C F A K R L I M S
 K T L R G Q A L G L V G L L V A L T K G
 AAACCTTGAGGGGTGAGGCGCTTGGACTGGTGGGCTTACTGGTACTTTAACCAGGGG 10.680
 621 TTTGGAAGTCCCGAGTCCGGAACCTGACACCGGAATGACCATCGAAATGGTTCGGG
 F R S P D P A Q V P P S V P L K L W P R
 A T S Q Q V I L T C L V V H Q T V Q A L
 CAACATCGCAACAGGTAATATTGTTGTTGGTGAACGAGCAAGGTTCAAGCTTTAT 10.740
 681 GTTGTAGCGTTGCTCATTAATTAAGAAACACCATGTTGGTCTGCCAAGTTGAAATA
 L M A V P L I L K N P P V G S P E L K I
 orf16 M P L E
 S E P T A E A Q K T T S Q E V S N A F R
 CGGAACCAACAGCGGAAGCAGAGAAACGAGCTCGCAGGAGTATCTAATGCTTTAGAA 10.800
 741 GGCTTGGTTGCGCTTCTGCTCTTTTGTGCGAGGCTCCTCCATAGATTACGGAATCTT
 R V L L P L M ←orf6

T R F S T S T T A N V E N F G T G V A P
 N T L F N L Y Y S O R R K L W Y R C S P
 ACACGCTTTTCAACCTCTACTACAGCCAACGTAGAAAATTTGGTACAGGTGAGCCCT 10.860
 10.801 TGTGCGAAAAGTTGGAGATGATGCGTTGCATCTTTTGAACCATGTCCACATOGGGA
 W T E A Y A N A W Q F S G D T D Y G Y M
 L D G S L C *
 TGGACGGAAGCTATGCTAATGCTTGGCAGTTCTCAGGGGACACAGACTACGGCTACATG 10.920
 10.861 ACCTGCTTCGGATACGATTACGAACCGTCAAGAGTCCCTGTGTCTGATGCGGATGTAC
 T N G N T T Y I Q Y G Q N D P S V W A S
 ACCAACGGGAACGACCTATATACAGTACGGGCAAAATGACCCGCTGTATGGGCGTCC 10.980
 10.921 TGGTTGCCCTTGTGCTGGATATATGTCATGCCCTTTTACTGGGAGACATACCCGAGG
 M R F W G E S V E I L E E T K N D D N S
 ATGGCTTTCTGGGGCGAATCGTTGAAATCCTAGAAGAGACAAAAACGATGATAATTCT 11.040
 10.981 TAOCGAAAGACCCCGCTTAGGCACTTTAGGATCTTCTGTTTTTGTCTACTATTAGA
 I T A K I R V K A L F W W S K R V S S N
 ATCAGGCTAAAAATCAGAGTTAAAGCCCTCTTTTGGTGGTCAAAACGGGTGAGTCAAAC 11.100
 11.041 TAGTGCGATTTTGTCTCAATTTCCGGAGAAAACACGAGTTTTCGCCAGTGGAGTTTG
 A G Y R V E Y D I K I N G R T V W T F S
 GCTGGTATCGGGTAGAATATGATATAAAATCAATGGGCGCACCGTTTGGAGCTTTAGC 11.160
 11.101 CGACCCATAGCCCTCTTATACTATAATTTTAGTTACCCGCGTGGCAAACTGCAATCG
 G Y T T D E V I K N D E V A Q D F T V T
 GGATATACACCGATGAAGTATTAAAAATGATGAAGTTGCCAAGACTTTACGTAACC 11.220
 11.161 CCTATATGTTGGCTACTTCAAGTAATTTTACTACTTCAACGGGTTCTGAAATGGCATTGG
 I P A E E S S S A S A L N I N V S Y P D
 ATACGAGCGGAAGAGTTTATCAGCCAGTGCCTTAAATATAATGATCTTATCCAGAC 11.280
 11.221 TATGTTGCGCTTCTTCAAGTAGTGGTCAAGGAATTTATATTACATAGAAATAGGCTG
 G Q Y S D N S F Y V G V Y L Y N T N K K
 GGGCAATATTAGACAAATCTTTCTATGTTGGTGTGCTTATATAATACCAACAGAAA 11.340
 11.281 CCGGTTATAGTCTGTGTANGAAGATACAAACACATGGATATATATGTTGTTCTTT
 K P K T L K P W A I R K S G I F K T L N
 AAGCTTAAAGCTTAAAGCGTGGGCAATCGTAAAGAGCGGAATCTTTAAGACCTTAAAC 11.400
 11.341 TTGGATTGTTGCAATTTCCGACCGGTTAGGCACTTCTGCGCTTAGAAATTTGGAATTTG
 R P S G I F Q O R K S S W O D I S E O P
 CGTCCCTCTGGAATCTTCCACAGCGCAAAAGCAGTTGGCAAGACATCAGCGAACGCCA 11.460
 11.401 GCAGGAGACCTTAGAAGTTGTGCGGTTTTCGTCACCGTCTCTGAGTGGCTTGTGGT
 A N A V G Q A V T A P H S V R K S G Q W
 GCAACGCGAGTGGGCAAGCAGTACAGCACCCACAGCGTGAAGAGTGGGCTCAATGG 11.520
 11.461 CGTTTGGCTCAGCCCGTTCGTCAGTGTGCTGGGTTGCGACTCTTTCAGCCGAGTTACC
 I G Q G Q I G Q E *
 ATAGGACAGGACAGATAGGACAAGAAATAGGGAGGTTTTCAGCCCTCTATTTTAAAG 11.580
 11.521 TATCTGTTCTGCTATCTCTGTTCTTATCTCTCCCAAGTCCGGAGGATAAAATTTTC
 orf17 M Q E S T K I W L Y A K S P F K N
 GAGAAGACATGCAAGAAATCAACCAAGATATGGCTTTATGCCAAAGCCCTTTTAAAGAG 11.640
 11.581 CTCTCTGCTAGCTTCTTGTGGTTCTATACGGAATACGCTTTTCCGGAATTTTTCG
 D Y A N V I N F E T R E S M E D F F T K
 ATTATGCCAATGTGATTAACCTTGAGACAAGGAAAGTATGGAAGACTTTTTCACAAAGA 11.700
 11.641 TAATACGGTTACACTAATGAAACTGTGTTCCCTTTTCACTCTGAAAAGTGTCTCT
 K N P H I E I V Y E Y D K F Q Y T Q R N
 AGAATCGCATATAGAAATTTGTGACGAGTATGACAGTTTCAATATACCCAAAGAAAGC 11.760
 11.701 TCTTAGGCGTATATCTTAAACACATGCTCATAGTTTCAAGTTATATGGGTTCTTTGC
 G S I V V S G R V E K Y E N V T Y M R F
 GCTCAATCGTAGTTTCTGGACGGGTAGAGAAGTATGAAATGTGACTTATATGCGCTTTA 11.820
 11.761 CGAGTTAGCATCAAGACCTGCCCATCTCTTCACTTTTCACTGAATATACGGGAAT
 I N N G R T Y Y A F V F D V L Y I N E D
 TCAACACGGAAGAACCTACTATGCTTTGTTTTGAGCTTCTTATATCAATGAAGATG 11.880
 11.821 AGTTGTTGCTTCTTGGATGATACGGAACAAAACTGCAAGAGATATAGTTACTTCTAC
 A T R I I Y E V D V W N T Y Q H E L K A
 CTACAGCATTTTACAGGTTGGAGCTTTGGAACACCTACGAGCAGCACTAAAGGCGC 11.940
 11.881 GATGTGGTAATAAATGCTCCACCTGCAACCTTGTGGATGGTGGTCTGATTTCCGGG

<p>LN V I G Q V E Q Q T L P N E L W A L K TTAATGTGATTGGTCAAGTGGAAACAGCAACCTTACCTAATGAGTTGTGGGGCTGAAAG 941 AATTACACTAACCAGTTACCTTGTGCTTTGGAAATGGATTACTCAACACCCGGCACTTTC</p>	<p>12.000</p>	<p>13.141</p>	<p>orf19 M L R L V A L Y S Q L L G G C CGGTGGGAGGGGCAATGTTACGGCTAGTGCCTTGTATTCTCAACTCTTGGGGCGTTG 13.200 GGCACCCCTTACCCTGTTACAATGCCGATCAGCGGAACATAAGAGTTGAGAACCCGCCAAC</p>
<p>DS Q Q G F S V G T K Y A T R A G E V G ACAGTCAGCAAGGTTTTTCAGTTGGGACTAATACGGCTACCGGGCTGGAGAGTTGGAA 001 TGTGAGTGGTCCAAAAAGTCAACCTGATTATATGOGATGTGCCGACCTCTCCAACTT</p>	<p>12.060</p>	<p>13.201</p>	<p>D I G G G I N T P N P N G N G R Q A S G CGACATCGGGGAGGTATCAACACCCCAACCCCAAGGCAACCGAGCAAGCGAGTGG 13.260 GCTGTAGCCCTCCATAGTTGTGGGGTTGGGGTGGCGTTGCGTCTGTTGCGTCAAC</p>
<p>ID T E W L V V A K P T I K H T T K A TGGATACGGAGTGGCTTGTAGTAGTGGCAAACTACGATTAAATGACCAAGGCA 361 AGCTATGCCTCACGGAACATCATCAGCGTTTGGATGCTAATTTACTGGTGGTTCGGTT</p>	<p>12.120</p>	<p>13.261</p>	<p>S D T Q R G H S A S N I G G D Y D W E A AAGCGATACACAAGGGGTATGAGTGCAGCAACATGGGGGCGACTAGTACTGGGAAGC 13.320 TTGCTATGTGTGCCCCATACTACGGTCTGTGAACCCCGCTGATACTGACCTTCG</p>
<p>NR P V N M S Y S G H Q K T F K Y F F I ACCGCCCCGTCAATATGAGTTATTGAGTATGCAAAAACCTTTAAATACTTTTTTATTC 121 TGGGGGGGAGTTTACTCAATAAGTCCATACGTTTTTTGGAAATTTATGAAAAATAAG</p>	<p>12.180</p>	<p>13.321</p>	<p>L G W K V R F D P S W A D L R V G C I V GCTGGGCTGGAAAGTCOGATTTGACCGCTCTGGGCTGATTACGGGTAGGGTGCTATTG 13.380 CGACCGACCTTTCAGGCTAACTGGGCGAAGCCGACTAAATGCCATCCGAGTAACA</p>
<p>P V N L K S G A S S R L F F R A K S M I CTGTAATTTGAAAAGCGAGCGTCCAGCGTTTATTTTCAGGCGAAAAAGTATGATAG 181 GACATTTAAACTTTTCGCTCGCAGGTGGCAAAATAAAGTCCGTTTTCATACTATC</p>	<p>12.240</p>	<p>13.381</p>	<p>C Y I P S G S N I W G H T S V I S A V N CTGCTATATCCAAAGCGGAAGCAATATCTGGGGCATACGTTCTGTTATTTACGGGTCAA 13.440 GACGATATAGGTTTGGCTTGTATAGACCCCGATGACGAGCAATAAGTGGCGAGTT</p>
<p>A F T L K T C I S T F S D * CTTTTACCTTGAAAACCTGTATAGCACCTTTTCGGATTGAATCAAGACCGCAGTAGCAC 241 GAAAAATGGAACCTTTGGACATATTCTGTGAAAAGCCTAACTTAGTTCTGCGCTCATGTG</p>	<p>12.300</p>	<p>13.441</p>	<p>G S S Y D V I E Q N Y A W N G Y T T E R CGGCTCAAGCTATGAGCTTATAGAACAAAATATGCTTGAATGGCTACACACGAGAACG 13.500 GCGAGTTGATGACTGCAATATCTGTTTAAATACGACCTTACGATGTGGTGTCTTGC</p>
<p>orf18 M Y L S R D I G V K Y K E T CGTAAACGAGATTGTGAACATGATTATTAAGCGGTGACATCGGGTAAATACAAAGAGAC 101 GCATTGGTCTAACACTTGTACATAAATCGGCAGTGTAGCCCATTTTATGTTTCTCTG</p>	<p>12.360</p>	<p>13.501</p>	<p>T G I D T I A N I E S I I Y P P E V V A CAGGGTATAGATACGATTGCCAATATTGAAGTATTATCTACCTCCGAGAGTTGTGCG 13.560 GTGCCATATCTATGCTAACGGTTATACTTTCATAATAGATGGGAGGCTTCAACAGCG</p>
<p>T E G D K T Y I E I L S N I T G S V A E AACCGAGGGGAGACAGCCTATATAGAAATCTGTCTAATATCAOGGAAGCGTTGACAGA 161 TTGGCTCCCGCTGTTCTGGATATATCTTTAGAACAGATTATAGTGCCTTCGCAAGCTCT</p>	<p>12.420</p>	<p>13.561</p>	<p>G G D I G E I T G N T G D R O L G N G D AGTGGCGATATTGGAGAGATAACAGGTAACACAGAGATAGACGTTAGGAAACGGGGA 13.620 TCCACCGCTATAACCTCTCTATTGTCCATTGTCTCTATCTGTCAATCCTTTGCCCT</p>
<p>I G R K N S R N Y R T S G S S S G G G S GATTGGCAGAAAGACAGCGAACTATCGGACATCTGGAAGTTTATCAGGCGGAGGAG 121 CTAACCGTCTTCTTGTGGCTTTGATAGCGCTGTAGACCTTCAAGTAGTCCGCTCCTTC</p>	<p>12.480</p>	<p>13.621</p>	<p>Y S K T A F D V E A L L I E V D V F F D CTACTCAAAACCGCTTTGACGTTGAAGCCTTACTAATTGAGGTGAGCTATTTTTTGA 13.680 GATGAGTTTTTGGCGAAACTGCAACTTCGGAATGATTAACTCCACCTGCATAAAAACT</p>
<p>G S T N E S G D I S T E E S R V R L V T TGGAAAGCAACGAAAGCGGAGACATTTCAACCGAGGAAGCGGTGTAGACTGGTTAC 181 ACCTTCGTGGTTGCTTTCGCTCTGTAAAGTTGGCTCTTTTCGACCAATCTGACCAATG</p>	<p>12.540</p>	<p>13.681</p>	<p>Y R P N V Y E I P N L L K I A Y D Q I Q CTATGCGCTTAAGTATATGAATCCGGAACCTCTTAAGATAGCTTATGACCGAGATACA 13.740 GATAGCGGATTACATATACTTTAGGGCTTGGAGGATTCTATGGAATAGTGGTCTATGT</p>
<p>R I I K K L V P D A T A E G I A G I I G TAGAATCATTAAAAAGCTAGTACAGACGACACAGCGGAGGTTATCGCTGGAATTTATCGG 141 ATCTTAGTAATTTTGTATCATGTTCTGGTGTGCGCTCCCATGACGACCTTAATAGCC</p>	<p>12.600</p>	<p>13.741</p>	<p>E G L R S Y M G K D D L E I E V Q L L N AGAGGGGCTAGGCTCATACATGGTAAAGACGACCTAGAAATAGAAGTACAATATTAAA 13.800 TCTCCCGATGCGAGTATGATACCATTTCTGCTGGATCTTTATCTTCTATGTTGATAATT</p>
<p>N F S A E S N V T A K K Y E A D Y A T G GAACCTTTGACGAGAAAGCAAGCTCACAGCAAGAAATACGAGGACGACTATGCTACTGG 101 CTGAAAAGTGTCTTTCGTTGCGAGTGTGGTTCTTTATGCTCGCTGATACGATGACC</p>	<p>12.660</p>	<p>13.801</p>	<p>S E F T E I E L Y D I Y G N S Y V Y Q P TAGTGAATTTACGAGATAGAGCTTTATGATATTACGGTAAAGCTTATGTTATCAACC 13.860 ATCACTCAATGCCTCTATCTCGAAATCTATAAATGCCATTGTCAATACACATAGTTGG</p>
<p>Y E Y E K M E S E P T A E N L M G S W G GTACGAGTACGAGAAATGGAATCAGAGCAACGAGCGGAGAACCTTATGGGAGCTGGGG 161 CATGCTCATGCTCTTTTACCTTAGTCTCGGTTGTGGCTCTTGAATAACCTTCGACCCC</p>	<p>12.720</p>	<p>13.861</p>	<p>Q Y L P R T I D E A H K Y K V I V S G S GCGATATTACGAGAACCATAGACGAGCACACAGTATAAGTTATTGTAAAGCGGAG 13.920 CGTCATAAATGGCTCTTGGTATCTGCTTGTGTGTTTCAATTTCAATAACATTGCGCTTC</p>
<p>A F A S L Y S I S L N E A G Y R G S D G CGCTTTTGCCAGTCTCTACTCTATCAGCTTAAATGAAGCTGGATATAGAGGTGAGACGG 21 GCGAAAACGCTCAGAGATGAGATAGTCAATTTACTTCGACCTATATCTCCAGTCTGCC</p>	<p>12.780</p>	<p>13.921</p>	<p>L G D S N Q V H I N F L E Y N N A N N V CCTTGGGGATAGCAATCAGGTTTCAATATCAATTTCTTGTAGTATAACACGCTAACATGT 13.980 GGAACCCCTATCGTTAGTCCAAGTATAGTTAAAGAACTCATTTGTTGOGATTGTTACA</p>
<p>N H W I G I G I G Q W T G P R A E E L L CAATCACTGGATAGGTATGGGATAGGACAGTGGACGGGCAAGGGCGGAGAGCTTTT 81 GTTAGTGACCTATCCATAACCTATCTCTGACCTGCGCTGGTTCCGCGCTCTCGAAAA</p>	<p>12.840</p>	<p>13.981</p>	<p>S Y A D K N I L D S L E S G D W A E H N AAGCTATGCTGATAGAATATTCTGGATAGCTTGGAAAGTGGCGACTGGGCGGAGACAA 14.040 TTGATAGCAGTATTCTTAAAGCTATCGAACCTTTACCGCTGACCGCGCTTGTGTT</p>
<p>N F A R S Q G K S L W D F N L Q F Q F M GAACCTTGCAGAAAGTCAAGGGAATCATTTGGGATTTTAACTACAATTTCAATTTAT 41 CTGAAACGTTCTTTCAGTTCCCTTTAGTAACACCTAAAATGGATGTTAAAGTTAAATA</p>	<p>12.900</p>	<p>13.981</p>	<p>P E H F K Y G L N D V T G K S V A I L N TCCAGAGCATTTTAAATACGGTTTGAATGACGTGACAGGAAAGCGTGGCTATTTTGAA 14.100 AGGCTCGTAAAAATTTATGCCAACTTACTGCTGCTCCCTTTTCGACCGGATAAAACT</p>
<p>N Q E S R A D T F R R V A S S T A S A S GAACCAAGAGAGCGGAGCGATAGCTTTAGACGGGTAGCCAGTTCCACAGCAAGCGCCAG 01 CTTGGTTCTCTCGGCTCGGCTATGCAAACTGCGCCATCGGTCAAGGTGTGTTTGGGGTC</p>	<p>12.960</p>	<p>14.041</p>	<p>W L M K F V T Q I V H C P F A H S N Q I D A E A S Y I Q S H K N Q M E H T Q L T 14.160 TGAGCGAGAGCGGCTATATCCAGGCCATAAGAATCAATGGAACATACACAGCTTAC</p>
<p>T N A S D F M N N W E G V A Y K E A E R CACCAATGCAAGCGACTTTATGAACAACTGGGAGGGGTAGCTATAAAGAGCGGAAGC 61 GTGGTTAGCTTCGCTGAAATACTTGTGTACCTCCCGCATCGGATATTTCTTCGCTTGC</p>	<p>13.020</p>	<p>14.101</p>	<p>A C T G G T C T T C G G T C G A T A G G T C T G G T A T T C T A G T T A C C T T G T A T G T G T G A A T G V C F G A I D L A M L I L H F M ←orf18</p>
<p>I E H * CATAGAGCACTAACGCTGTTATCTACGATACAAGCAACTACAGAAAGGTAACAA 21 GTATCTGCTGATTGGGACCAATAGATGCTATGTTCTGCTGATGCTTTTCCCATTTGTT</p>	<p>13.080</p>	<p>14.161</p>	<p>F K E N R D M L K Q S V D L S N K Q V A GTTTAAAGAGAAATCGGACATGCTGAAACAGAGCGTAGACCTATCTAATAACAGTGGC 14.220 CAAAATTTCTCTAGCCCTGACGACTTTGCTCGCATCTGGATAGATTATTGTTTCAGCG</p>
<p>TGGCAGAGCAACAGAAACGCTAAAGCACTAATGAATCAAGTACCGGTAGGAACAA 81 ACCGTTCTGTTGCTTTCGATTTTCGATTTTACTTTAGTTTCACTGCCCATCTGTT</p>	<p>13.140</p>	<p>14.221</p>	<p>T A N S Q A S Y N A Q F A V D S A N I N TACAGCAACTCACAGCCAGCTATAACGCAAAATTTGCGGTAGATAGCGCAATATCAA 14.280 ATGTCGTTGAGTGTTCGGTGTATGCTGTTTAAACCGCATCTATCGGGTTATAGTT</p>

Q W T E G A S G I L N V A G N L L T G N
CCAATGGACGGGGCGCTAGTGGTATCTTAAAGTAGCTGGAATCTCTTAACAGGGAA 14.340
81 GGTTACCTGCCCTCCGCGATCACCATAGAAATTTGCATCGACCTTTAGAGAAATGTCCCTT
F G G A L G G L A S G G M K V F N A N R
CTTTGGGGGCGCACTTGGTGGCTTGGCTTGGTGGTATGAAAGTCTTAAAGCTTAACCG 14.400
41 GAAACCCCGCGGTGAACCCCGAAGCGACACCACATACTTTGAGAAATTCGATTGGC
D Y N D K V V Q Q G F T S E N N A L K S
TGATTATAATGATAAAGTGGTGCACAAAGGTTTACAAAGCGAAACAAAGCTTTGAAATC 14.460
01 ACTAATATTACTATTTTACCACGTTGTTCCAAAGTGTTCGCTTTGTTGGGAACTTTAG
Q S N A L A N M K S K I A L D Q S I R A
TCAGTCAACCGCACTTGCACACATGAAATCTAAGATAGCACTTGACCACTTATCAGAGC 14.520
61 AGTCAGTTTGGTGAACGTTTGTACTTTAGATTCTATCGTGAAGTGGTCAGATAGTCTCG
Y N A T M A D L O N O P I S V Q Q I G N
TTACAACGCAAGTGGCAGAGCTACAGAACCGCTTACGCTGCAACAAATCGGGAA 14.580
21 AATGTTGGCTGGTACCGCTCGGATGCTTGGTGGATAGTGGCAGTTGTTTAGCCCTT
D L A F Q S G N R L T D V Y W K V S L A
TGACCTTGCTTTCCAGAGTGGGACAGAGTACAGAGGTTTATGGAAGTCTCCCTAGC 14.640
81 ACTGGAACGAAAGTCTCACCTTGTCTGACTGTCTGCAATAACCTTTAGAGGGATCG
Q K E I M G R A N E Y I K C Y G V L V N
TCAAAAGAAATCATGGGACGGGCAAGAGTACATAAAATGCTATGGGGTCTTGTCAA 14.700
41 AGTTTTCTTTAGTACCTCGCCGCTTTGCTCATGTATTTTACGATACCCCAAGACAGTT
W F T N D A L S V M R S R K R F N Y I K
CTGGTTCATAATGACGCTTAAGTGTAATGAGGTCAAGAAACGCTTTAATATATCAA 14.760
01 GACCAAGTGATTACTGCGAGATTACATTACTCCAGTTCTTTTGGGAAATTAATATAGTT
M I N V N L G T L R A N O S H M N A I Q
GATGATTAAAGTTAACCTTGGACACTAAGAGCAACAGTGGCAGATGAACGCTATTCA 14.820
61 CTACTAATTGCAATTTGAACCTTGTGATTCTCGGTTGGTCAGGTTGACTTGGGATAAGT
A I F Q S G V R I W N Y S A N K E D G I
GGCTATTTTTCAGTCAGGTGTGAGAAATTTTAAAGTATGATATGAAATAGAA 14.880
21 CCGATAAAAGTCAGTCCACAGTCTTATACCTTAATAAGTGGTTATTTCTTCTCGCGTA
L F D I Q K N N P N F *
TTTGTGTTGATATTGAGAAACACCGCAATTTTAAAGTATGATATGAAATAGAA 14.940
81 AAACAACTATAAGTCTTTTGTGGGCTTAAAAATTTTCACTATATATTCTTTATCTT
orf20 M E A T E K W Y N P Q K H L S
AGGAGTGTATTTCTATGGAAGCACTGAAAAATGGTAAATTCGAGAAAAATGCTATCTT 15.000
41 TCCTCACTAAAGATACCTTGGTGTACTTTTACCATGTTAGCGCTCTTTAAGATAGAA
Y N Q Y L N F V I G G R G I G K T F A L
ATAACAGTATTTAAACTTTGTTATCGTGGTGGTGAATTTGTAAGACCTTTGCACTCA 15.060
01 TATTGGTCATAAATTTGAAACAATAGCCACGACCACTTAACGATTCTGGAACGAGT
K K Y L F K R F I E K G E Q F I Y L R R
AGAAATATCTGTTTAAAGGTTTATTGAGAAAGGGGAACAATCTACCTAAGCGGA 15.120
61 TCTTTATAGACAAATTTTCCAAGTAACTCTTCCCTTGTTAAGTAGATGGATTGCGCT
N K S E L D R I D K D K F F T T E L L K
ACAAGTCAGAGCTTGACGCTATGCAAGGCAAGTTTTACTACGAGTTATTAAAGC 15.180
21 TGTTCACTCTGCAAGTGGCATAGCTGTTCTGTTTCAAAAAATGATGGCTCAATAATTTG
Q V F T N F E V I D S D A S K I H T K I
AAGTTTCTACTAATTTGAAGTATTGACAGCGACCGCAAAATCCATACTAAGATTA 15.240
81 TTCAAAGTGATTGAACTTCACTAAGTGTGCTGGCGTGGTTTAGGTATGATTCTAAT
I F R A D N M E E E E N I L V L S S T K
TTTTCAGAGCGGACCAATGGAGGAGGAGGAAATATCTGTTTATCTTCTACTAAAA 15.300
41 AAAAGTCTGCGCTTGTGATCTCTCTCTCTTTATATGAACAAATAGAAATGATTTT
I I L N G K I V C Y L K S L S T W V D L
TTATTCTTAACGGGAAATCGTTTGTATCTCAGAGCGCTTTCTACATGGGTAGACTTGA 15.360
01 AATAAGAATGGCTTTTAGCAACAGATAGAGTTCTGGGAAAGATGTACCATCTGAAT
K G S E Y D E V M S I L Y D E V L I D V
AAGGTGACAGATGATGAGGTAAATGATGATTCTATACGATGAAGTATTGATAGAGTTA 15.420
61 TTCCAGTCTCATACTACTCCATTACTCATAGATATGCTACTTCACTAATATCTGCAAT
T S K K R Y L D N E V E A L L N F I F S
CCAGTAAAGAGGATCTCGATAACGAGGTGGAAGGTTACTAAATTTCTATCTTCTCAG 15.480
21 GGTCATTTTCTCCATAGAGCTATTGCTCCACCTTGGCAATGATTAAAGTAGAAGATC

V F R R R D G C H A Y L L S N A S N F N
TATTCAAGAACGGGACGGGTGTCATGCTTATCTACTATCAATGCAAGTAAATTTCAACA 15.540
15.481 ATAAGTCTTCTGCCCTGCCACAGTACGAATAGATGATGTTTACGTTTCAATAAAGTTGT
N P Y F A F L K F Y D D N G K R F Y N L
ACCCTTATTTTGGCTTCTCAAGTCTATGACGACAAAGGCAAGCGCTTTTCAACATCTGA 15.600
15.541 TGGGAATAAAAGCGAAAGAGTTCAAGATACTGCTGTTCGCGTTGCGGAAATGTTAGACT
K Q Y A T L I E F P P H S A F Q T E E E
AACAAACGCAACCTTGATAGAGTTCGCCCTCATTCAGCCTTTCAACAGAGGAAGAAA 15.660
15.601 TTGTTATGCGTTGGAACATCTCAAGGGGGAGTAAAGTGGGAAAGTTTGTCTCTCTTT
K E S G F F K L L S K S S I Y E S V A N
AAGAGAGCGGATCTTTAAGCTATTGAGTAAAGTCAAGTATTATGAAGCGTTGCTAATA 15.720
15.661 TTCTCTGCGCTAAGAAATTCGATAACTCATTGAGTTTCAATAACTTTTCCAGCAGATTAT
N E F Q I K N D K N I A K I K G L K S R
ATGAATTTGAGTAAAGAACGACGAAGTATAGCAAAATTAAGGCTTGAAGTCTAGGC 15.780
15.721 TACTTAAAGTCTAATTTCTGCTGTTCTTATATCGTTTTTAAATTTCCGAACCTTCAGATCG
L Y S F Y C D G T F L T G Y Y I D N M V
TATATAGTTTCTATTGTGACGCTACTTTCTTAACGGGGTACTATATCGCAATATGGTAT 15.840
15.781 ATATATCAAGATAACACTGCCATGAAGAAATGGCCCATGATATAGCTGTTATACCATA
Y I A K G F D K N L T A Y C L E V E Q V
ATATGCTAAAGGCTTTGACAAAAATTTGACCGCTTATGTTTAGAAGTGAAGCAGGTGG 15.900
15.841 TATAAGGATTTCCGAACTGTTTTTAACTGGGGAATAACAAATCTTCACTTGTCCACC
E D G F V Y L N K A S A L G K T L R S L
AAGAGCGGTTTGTGACTTGAATAAAGCCAGCGCACTAGGTAAACTTTAGGAGTTTGT 15.960
15.901 TTCTGCGCAACACATGAACCTTATTTCCGTCGCGTATCCATTTGAAATGCTTCAACA
Y L K N M F I Y E D L E T K N K F I E V
ACTTAAAAATATGTTTATTATGAAGATTAGAAACTAAAAACAAATTTATAGAGTTA 16.020
15.961 TGAATTTTTTATACAAATAAATCTTAAATCTTTGATTTTTTGTAAATATCTCCAAT
orf21 M L Y N I M L E V A K G D Y I T I L
I N H V I *
TCAATCATGTTATATAATATTATGCTAGAAGTTGCTAAAGCGACTATATCACAAITCTT 16.080
16.021 AGTTAGTACAATATATTATAATACGATCTTCAAGATTTCCGCTGATATAGTTTAAAGAA
F A L I L F D F I T G F L K A W K W K V
TTTGTCTTGTATCTGTTGACTTTATCACAGGCTTTTAAAGGCTTGAAGTGAAGAGTG 16.140
16.081 AAACGAACTAAGCAAACTGAAATAGTGTCCGAAAAATTTCCGAACTTCACTTTTCA
T D S W T G L K G V I K H T L T F I F Y
ACCGATAGCTGGACAGGGTTAAAGGAGTTATCAACATACCCCTAACATTTATTTTTTAC 16.200
16.141 TGGCTATGACCTGTCCCAATTTTCTCAATAGTTTGTATGSGATTTGAAATAAAAATG
Y F V A V F L T Y I H A H A V G Q I L L
TATTTGTAGCGGTATTCTAACATATATTCAAGCTATGGCAGTGGTCAAGTTTGTCTT 16.260
16.201 ATAAACATCGCCATAAGAAATGTATATAAGTGCGATACGTCAGCCAGTCTAAAAAGAA
V I I N L Y Y A L S I M E N L A V M G V
GTTATCATTAAGTGTACTATGCACTTTCAATATGGAATACTTGTGTTATGGGTGTC 16.320
16.261 CAATAGTAATTGAACATGATACGTGAAGTTAATACCTTTTGAACGACAAATACCCACAG
F I P K F M T A R V Q E E L Q K Y T A Q
TTTATCCCTAAATTTATGACGGCTAGAGTACAGGAAGATTCGAGAAATACACCGCACAA 16.380
16.321 AAATAGGATTTAAATACTGCGGATCTCATGTCTCTTAAAGCTTTTATGTTGGGTGTT
orf22 M V K
L D A G K D L L E E F K G E K K *
CTAGACGCGGAAAGACCTGCTAGAAGAAATTTAAAGGCAAGAAATATGGTTAAAA 16.440
16.381 GATCTGCGCTCTTTTCTGGAAGATCTTCTTAAATTTCTCTTTCTTTATACCAATTTT
K N D L F V D V S S H N G Y D I T G I L
AGAATGATTATTGTTAGATGTTTCAAGTCAACAGGTTACGATATAACAGGATCTTTGG 16.500
16.441 TCTTACTAAATAACATCTACAAAGTTCAGTGTGCAATGCTATATTGCTCATAGAACC
E Q M G T T N T I I K I S E S T T Y L N
AGCAATGGGAACAACTAACACCATCATTAATTTCTGAAAGTACGACCTATTAAACCC 16.560
15.501 TCGTTTACCCTTGTGATTGTTGAGTAAATTTTAAAGACTTTTATGCTGGATAAAATTTGG
P C L S A Q V E Q S N P I G F Y H F A R
CTTGCTTGTCTGCTCAAGTGGAGCAGTCAACCCCTATTGCTTTTATCACTTGCACGCT 16.620
15.561 GAACGAACAGACAGGTTTCACTCGTCAGTTTGGGATAACCGAAAAATGTAAGCGTGGCA

F G G D V A E A E R E A Q F F L D N V P
 TTGGCGGAGAGCTAGCAGAAAGCGAAGAGAGCGAGTTTTCCTTGCAACGGTCCCTA
 .621 AACCGCCTCTGCATCGTCTTCGGCTTCTCTTCGGCTCAAAAAGGAAGTGTTCACGGAT 16.680
 M Q V K Y L V L D Y E D D P S G D A Q A
 TGCAAGTTAAATACCTGTATTTGGACTACGAGGAGACCGCAAGCGGAGAGCGCAAGGGA
 .681 ACGTTCATTTATGGAACATAACCTGATGCTCTGCTGGGTTCGGCTCTGGTGTTCGCT 16.740
 N T N A C L R F M Q M I A D A G Y K F I
 ACATAACGCATGCTTACGCTTTATGCGATGATTGCTGACGCTGGATATAAACCCTATTT
 .741 TGTGATTGCGTACGAATCGGAATACGTCTACTAACGACTGCGACCTATATTTGGATAAAA 16.800
 Y Y S Y K P F T H D N V D Y Q Q I L A Q
 ATTATAGTTATAAACCGTTTACACATGATAATGTGGACTATCAGCAATCCTTGACAGT
 .801 TAATATCAATATTTGGCAATGTGTACTATTACACCTGATAGTGTGTTAGGAACGTGTCA 16.860
 F P N S L W I A G Y G L N D G T A N F E
 TCCCTAATTTCTATGATTGCGAGCTATGGCTTAAACGATGTACAGCTAATCTTGAAT
 .861 AGGGATTAAGAGATACCTAACGTCGATACCGAATTTGCTACCATGCGATTGAAACTTA 16.920
 Y F P S M D G I R W W Q Y S S N P F D K
 ACTTCCCAAGCATGGACGGGATAAGATGGTGGCAGTATTCTAGTAACCGGTTTGCAAGA
 .921 TGAAGGGTTCGATACCTGCCCTATTCTACCCCGTCATAGATCATTGGGCAAACTGTTCT 16.980
 N I V L L D D E E D D K P K T A G T W K
 ATATTGTACTGTTAGACGATGAAGAGACGACAGCAAGCAAGACCGCTGGAACTGGAAAC
 .981 TATAACATGACAATCTGCTACTTCTCTGCTGTTGGTTCCTGGGACCTTGACCTTTG 17.040
 Q D S K G W W F R R N N G S F P Y N K W
 AAGACAGCAAGGGTGGTGGTTCAGACGAAACATGGCAGTTTCCCTTATAATAATGGG
 .041 TTCTGTGTTTCCCCACCACCAAGTCTGCTTTGTTACCGTCAAGGGAATATTATTTACCC 17.100
 E K I G G V W Y Y F D S K G Y C L T S E
 AAAAAATCGGTGGTGTGGTACTACTTCGATAGTAAGAGATATTGCTTAAAGAGCGAAT
 .101 TTTTATGACCACACACCATGATGAAGCTATCATTCTCTATAAGAAATGCTCGCTTA 17.160
 W L K D N E K W Y Y L K D N G A M A T G
 GGCTCAAGATAATGAAAAATGGTACTACCTCAAGGACACCGCGCAATGGGCGACTGGT
 .161 CCGAGTTCTATTACTTTTACCATGATGGAGTTCCTGTTGCCGGTTCACGCTGACCAA 17.220
 W V L V G S E W Y Y M D D S G A M V T G
 GGGTCAAGTACGGGTGAGAGTGGTATTATATGGAAGATTGAGGCGCTATGGTTACTGGT
 .221 CCCAGATCAGCCAGTCTCACCATAATATACCTGCTAAGTCCGCGATACCAATGACCAA 17.280
 W V K Y K N N W Y Y M T N E R G N M V S
 GGGTCAAGTATAAGAAATACTGGTACTATATGACAAATGAACGCTGGTAACTGTTCTA
 .281 CCCAGTTTCAATTTCTATTGACCATGATATACTGTTTACTTGACCAATTGTACCAAGAT 17.340
 N E F I K S G K G W Y F H N T N G E L A
 ATGAATTTATAAGTCTGAAAAGGTTGGTATTTCATGAACACAAAGGAGGCTGCGAG
 .341 TACTTAAATAATTGACACCTTTTCCAACATAAAGTACTTGTGTTTGCCTCTCGAACGTC 17.400
 D N P S F T K E P D G L I T V A +
 ACAATCCAAGTTTTCAGAAAGAACGAGCGGCTTATAACGCTAGCATAAAAAAGAAAAG
 .401 TGTAGGTTCAAAGTCTTTCTTGGTCTGCCGAAATATTGGCATGTTATTTTCTTTTC 17.460
 CTAGTAGGATTTTCTACTAGCTGTTTTATAGTCTGCTATAATTTTATAAGCATCTTCG
 .461 GATCATCTCTAAAAGGATGATCGACAAAATATCAGACGATATTAATATTCGTAGAAGC 17.520
 TCTGGATTATCCAGAGCGATGGAACAGATTGCGACAGGACGCTGTTCACTGATTGTAT
 .521 AGACCTAATAGGTCTCGCTACCTTGTCTAACGCTCTGCGACAGTATGACTAACATA 17.580
 TCTGTAGTAGTGATTTTCTAGCTGTTTGTAAITGCTGATATGCTTTTCAATCCAGCC
 .581 AAGACATTCATCACTAAAAGATCGACAAACATTAAAGACTATACGAAAGTATAGGTCG 17.640
 AGAGAAAAGGAGTGGTGTAGGTTTATGAGTGTGGTGGTCACTTGTCTGTCTAGCCG
 .641 TCTCTTTTCTCACCACATCCAAATCTCAACGACCCAGTCAGAACGSAACAGATCGGGC 17.700
 TCAGGTATGCTTTTCACTTTCTGAAGCGTTTCTAAATAGCTTGAATGAATGTATAA
 .701 AGTCCATACGAAAGTAGTAAGAGCTTTTCCAAAGATTTATCGAACATTTACTTACATATT 17.760
 CCAGTTTCATCAAGCGTTAOCATTCCCGAAGCGTGAAGCGTTAGGCTTGAATCTTTG
 .761 GGTCAAGTAGTTTCGAATGGTAAGGGCTTGGACCTTCGAATCGAACGTTAGAAAAC 17.820
 ACCGTGTAACCTTAAACTGTTTCTAAATTTGAAGGCTATTCTTGTGTTTCTAAGCGTGT
 .821 TGGCACATTGGAATTTGACAAAAGTTTAACTTCOGATAAAGACGACAAAGATTGCGACA 17.880

TTTGCTTCCATAGTATTAACCCCTACGTTTTCAGATAAGCGACTAGACGCAAAATAG 17.881 17.940
 AAACAGAGGTATCATAATTGGGGGATGCAAAAGTTCTATTGCTGATCTGGGTTTATC
 * G G V N E L Y A V L R L I S
 AAGCGGTGTACTACTGTTTTCTCGCTTTAACTTCGTGATGGTCTTTGTTGTTGCTCAA 17.941 18.000
 TTGGGCACATATAGCAAAAGAGCGAAATGAAGCACTACCAAGCAACCAACGAGTT
 A T N S D N E R K L K T I T S Q Q Q E L
 GTTGTTTTTGTTGCTTGAAATGGTATAGGTCAAGAGTGACATAATCACTAGAAAAACCA 18.001 18.060
 CAACAAAACAACGAACCTTTTACCATATCCAGTTTCTCACTGTATTAGTGATCTTTTGGT
 Q K Q Q K F I T Y T L L S M I V L F V L
 AGATGGCAAGGTTACAAGAAGCGAACCCAAAACAGAACGACCTTTTTTGTTTAAT 18.061 18.120
 TCTACGTTTCAATGTTCTTCGCTTGGTGGTTTTGGTCTTTTGGTGGAAAAACAATTA
 I A L N C S A F W W F W F R G K K N L K
 TGTATATGAATTTTTCATTTTCTAATACCTCAATTTGAATAATCTCTTTCTACTGAT 18.121 18.180
 ACAATATACTTAAAGAGTAAAGAAATTTATGGAGTTAACTTTATTAAGAGAAAGTAGCTA
 N Y S N K M ←orfα
 AAATAAGTAACCTCTAATATACCTTTGTAAGCTGTAGACCCCGTCTGGTTGGTAATGTTG 18.181 18.240
 TTTATTCTTGGAGATTATATGGAACATTTCGACATCTGGGCGAGCAACCACTTACAAAC
 TTCTACATCATATAAGTTATAAGTAGTGGCTCAAATGCCCTCTATATCTGTCATTATAGG 18.241 18.300
 AAGATGTAGTATATTCAATATTTCATACCGAGTTTACCGGAGATATAGCAGGTAATATCC
 CTTTATACGACTAATAGCATAATCTAAGGAAATGATAGGAAAGCCATCAAAATAATTTG 18.301 18.360
 GAAATATGCTGATTATCGTATTAGATTCTTTTACTATCTTTTCGTTAGTTTATTTATAAC
 CTTGACGACCTTTCTAACCGCTTCTATTTTCAGTTGGAAAAATCTGTTTAAATCGTTCAT 18.361 18.420
 GAAGTCTGGAAGAGTTGGGCAAGATAAAGTCAACCTTTTGAAGCAAAATTTAGCAAGTA
 CTAGTAACCGTCCCTTCTTGTACATAGCGCTCAAGCGTGTATTTTAAACAATAATGTGA 18.421 18.480
 GATCATTGGCAGGGAAGGAACATGATCGCGAGTTTGGGACTAAAATTTGTTATTACACT
 * Y G D R G Q V Y R E F A S K L C Y H S -
 ATATTTTGAAGTATTTCAAGTTTTCCTTCTCAAGATAATAAGCGTGTTTTAAAGCGCTC 18.481 18.540
 TATAAACTCATAAAGTTCAAAAGTGAAGGAGTTTCTATTATCGCACAAATTCGCGAG
 Y K S Y K L N K V E E F I I L T N L R E -
 orf23 M D F F I S T I S I Y L F F
 ATTTTAAATAGTTCTCATGGACTTCTTTATATCCAGATTCTATATACCTCTTTTTC 18.541 18.600
 TAAAAATTTTATCAAGAGTACCTGAAGAAATATAGGTGCTAAAGATATATGGAGAAAAG
 N K Y N E H V E K Y G R N R Y V E K E
 C F Q F I V W K L I N T L Y F K R S Y P
 TGTTCAGGTTTATCGTCTGGAACTTATCAATACATTGTATTTTAAAGAGGATTACCCC 18.601 18.660
 ACAAGGTCAATAGCAGACCTTTGAATAGTTATGTAAACATAAAATCTCTCAATGGGG
 T E L K D D P F K D I C Q I K L P T V G
 L K N N L K L C Y H V Y P P I C T N T R
 CTCAGAAATAATCTGAAATTTGCTATCATGTTTACCCTCTATTTGTACAAATACAGCT 18.661 18.720
 GAGTTCCTATTAGACTTTAACAGCATAGTACAAATGGGAGGATAAACATGTTTATGTGCA
 E L I I Q F Q A I M ←orfβ
 F R F H N I Q H I L N Q N K I G N F Q N
 TTTGTTTCCCAATATCCACATATTTCTGAACCAAAATAAGTAGGAACTTTCAAAAT 18.721 18.780
 AAAGCAAGGTTTATAGGTTGTATAAGACTTGGTTTTATTCTATCTTTTGAAGTTTAA
 F Q R S K A S F V P S T *
 TTTCAAGCTCCAAGGCTTCTTTGTACCTCTCATATAAGAGTATAACCCCTTTTCTCT 18.781 18.840
 AAAGTTGCGAGGTTCCGAAGAAACATGGGAGATGTTATCTCATATTGGGAAAAAGAA
 * R E L A E K T G E V Y S Y L G K E E
 AGGTTTGTCTTCCCTACAAAAATACAGGTGTTTCTATCTGTAAACCTTGACTTTTAGCA 18.841 18.900
 TCCAAACGAAAGGATGTTTTATGTCACAAAGATAGACATTTTGAAGTGAATTCGT
 L N A K G V F I C T N R D T F G Q S K A
 AGTTTCACTAATTTATCTTCAITGTTATGCGCTGCTGCATATGATCATCACTAGTTA 18.901 18.960
 TCAAAGTGATTAATAGAAAGTACAAGTACGAGCAAGTATCAGTAGTGTATGACCAAT
 L K V L K D K M N M G A Q M T M V Y Q N
 TTGATAATCATTTTGTATACCTCTTTATCTTCTGATACCATATAGCATTTTCTCAAA 18.961 19.020
 AACTATTAGTAAAAAATATGGAGAAATAGAAAGACTATGGTAAATATCGTAAGAGTTTT
 N I I M ←orfα
 TGCTTGCAAGTGTTTTTTGAATTTATTTTAAATTTATTTTGAAGATTTTTTATCTCA 19.021 19.080
 ACGAAGCTTCACAAAAAATTTAATAAAATTTAATAAAATCTTTACAAAAATAGAGT

```

081  TTTCTAATATATACATTGTATCATTTTATAAAATGATTGTCAAGAGATTTTGAATTA 19.140
      ↓
      AAAGATTGATATATGTAAAGATAGTAAATATTTTACTAAGATTCTCTAAAACTTTAAT

141  TTTTATTTTATTAGTGATTTTATTCTTGACATTTTCTAAATTGTGATATAATGATA 19.200
      AAAAAATAAATAAATCACTAAATAAAGAACTGTAAAAAGATTTAACACTATATTACTAT

201  ATAGAAACAGATGAAGGATTTTAGAAGATGATAGATGATTGATTTTGAATGGCTAGT 19.260
      TATCTTTGTCTACTTTCTTAAATCTTCTACTATCTACTAATAAATCTTACOGATCA

261  GTTGGAAATGGATTTTCTCTTTCATTTCATTTTCAAATTTCCATTGGGGGAATTTTGAA 19.320
      CAACCTTTACCTAAAGAGAAAGTAAAGTAAAGTTTAAAGGTAAACCCCTTAAAACTT

321  AAAGGGGTGGGGAGTACATGCTTT 19.345
      TTTCCCAACCCCTCATGTAAGAAA

```

figura 4. Secuencia de nucleótidos del DNA de Cp-1. Se indica la doble cadena de todo el genoma y la secuencia de aminoácidos deducida de las ORFs, en código de una letra. Los asteriscos representan los codones de terminación. Los BS están subrayados, las cajas -10 y -35 de las regiones promotoras aparecen doblemente subrayadas y las de los promotores no funcionales que se dirigen hacia los extremos aparecen con subrayado discontinuo. Los sitios de inicio de la transcripción están marcados con flechas verticales. Los presuntos terminadores de la transcripción están señalados con flechas horizontales. Los nucleótidos que forman las ITRs en los extremos del genoma aparecen en negrita. La secuencia representada aparece en la base de datos GenBank/EMBL con el número de acceso Z47794.

Tabla 6. Codones de iniciación y de terminación de las ORFs de Cp-1 y sus correspondientes masas moleculares

ORF	Fase	Secuencia ^a	Ventana ^b	Codón de iniciación (nt)	Codón de terminación (nt)	Nº de aminoácidos	Tamaño (kDa)	Comentario
<i>CADENA SUPERIOR</i>								
1	1	TAAAC AAAGGAG AAAGAAACATG	10	ATG (376)	TAA (640)	88	9,7	
2	3	TTGAGT AAAGGAG ATAATAAAATG	10	ATG (657)	TAA (945)	96	10,6	
3	3	TACAAA ATGAAGAT GGGAATGTG	9	GTG (1.152)	TAA (1.344)	64	7,1	
4	1	CAAGCG GAAGGGT GTAATTAAATG	10	ATG (1.351)	TGA (2.041)	230	26,8	Proteína terminal
5	3	AAGAC GAAGAAG CGAGGAAGAAGTG	12	GTG (2.040)	TGA (3.744)	568	66,2	DNA polimerasa
6	1	CAGATT AAAGGGG TACGCTGATTTATG	13	ATG (3.703)	TAA (4.147)	148	17,6	
7	2	TCAACA ACGGAG ATAAACAAATG	10	ATG (4.142)	TAA (4.652)	170	18,8	
8	1	AAATT GAAAGG AGAGGGCTATG	7	ATG (4.876)	TAA (5.164)	96	10,6	Proteína de andamiaje
9	1	TAGAAG AGGGAG AAGAATAGAATG	10	ATG (5.410)	TAA (6.505)	365	42,1	Proteína de la cabeza
10	3	TAAGT GAAAGG GAAGTAATG	6	ATG (6.564)	TAA (7.575)	337	39,5	Conector
11	1	CCAGAT ATAGAA ATCCCGTG	5	GTG (7.627)	TAG (8.209)	194	22,6	Collar
12	1	TTTAG AAAGGT GTATAGATG	6	ATG (8.224)	TGA (8.806)	194	22,7	
13	3	CAGCAG AAAGAG GACATGAAAACCTATG	13	ATG (8.805)	TAA (9.027)	74	8,4	
14	3	TAAAG AAAAGG GTAGAAACAAATG	11	ATG (9.105)	TAA (9.990)	295	32,9	
15	3	AATAAG AAAGG ATTTTAAACATG	9	ATG (10.014)	TAA (10.878)	288	33,2	
16	1	GACCTCG CAAGG ATCTAATG	8	ATG (10.789)	TAA (11.548)	253	29,2	
17	3	ATTTTT AAAGG AGAAGACATG	6	ATG (11.589)	TGA (12.279)	230	26,2	Cola
18	2	CACCGT AAACC AGATTGTGAACATG	10	ATG (12.320)	TAA (13.031)	237	26,4	
19	2	CCGTGGG AAAGT GGGCAATG	4	ATG (13.157)	TAA (14.915)	586	64,9	Cola
20	3	AAATAG AAAGG AGTGATTTTCTATG	10	ATG (14.955)	TAA (16.035)	360	42,2	Encapsidación
21	1	AAATTT ATAGAG TTATCAATCATG	10	ATG (16.027)	TAA (16.429)	134	15,4	Holina
22	3	GAATTT AAAGG AGAAAAGAAATAATG	11	ATG (16.431)	TAA (17.448)	339	39,3	Lisozima
23	1	ATTTTT AAAAT AGTTCTCATG	6	ATG (18.559)	TAA (18.817)	86	10,4	
<i>CADENA INFERIOR</i>								
a	2	AAAGAT AAAGG GTATAACAAATG	10	ATG (18.971)	TAG (18.785)	62	7,0	
b	3	GTACAA ATAGG AGGGTAAACATG	8	ATG (18.690)	TAG (18.423)	89	10,7	
c	1	TATTCA AAATTG AGGTATTAAGAAAATG	12	ATG (18.139)	TAA (17.899)	80	8,8	
d	1	CTCTTT AAACG TAAGCTGTGTATG	9	ATG (14.149)	TAA (13.870)	93	10,2	
e	3	CTCCTG CGAGG TCGTTTCTGTG	8	GTG (10.761)	TGA (10.554)	69	7,6	
f	3	TGGTAG AAATAG TTGTAGGTCATG	9	ATG (6.419)	TGA (5.978)	147	16,2	
rRNA 16S ^c		3' UUUCUC						

^a Secuencias de RBS de las ORFs de Cp-1 alineadas con la secuencia complementaria del rRNA 16S. En cada ORF, el RBS se muestra en **negrita** y se subraya el codón de iniciación.

^b La ventana es el número de pb entre el codón de iniciación y el nt del RBS que puede emparejar con la U en la secuencia CCUC del rRNA 16S.

^c Secuencia del extremo 3' del rRNA de 16S de *S. pneumoniae* (Bacot y Reeves, 1991).

ORF15-16, y ORF20-21, el solapamiento es de 44, 8, 92 y 11 nt respectivamente (Figura 4, Tabla 6).

La distancia media entre los RBS y el codón de iniciación es de 9 nt para las 29 ORFs identificadas en ambas cadenas (el rango varía de 4 a 12), lo que se corresponde bien con la distancia encontrada para los genes procariotas (Hager y Rabinowitz, 1985). Por otra parte, se han localizado, usando el programa PCFOLD (Zuker y Stiegler, 1981), varias posibles estructuras de horquillas de terminación de la transcripción localizadas en el extremo 3' de las ORF5, 7, 9, 13, 16, 18, 19, 22 y c (Figura 5). Conviene resaltar que los organismos con contenido alto en A + T en su DNA, como neumococo, presentan pocos pares G + C en estas estructuras, por lo que el ΔG no alcanza valores tan negativos como los postulados para *E. coli* (Carafa y cols., 1990).

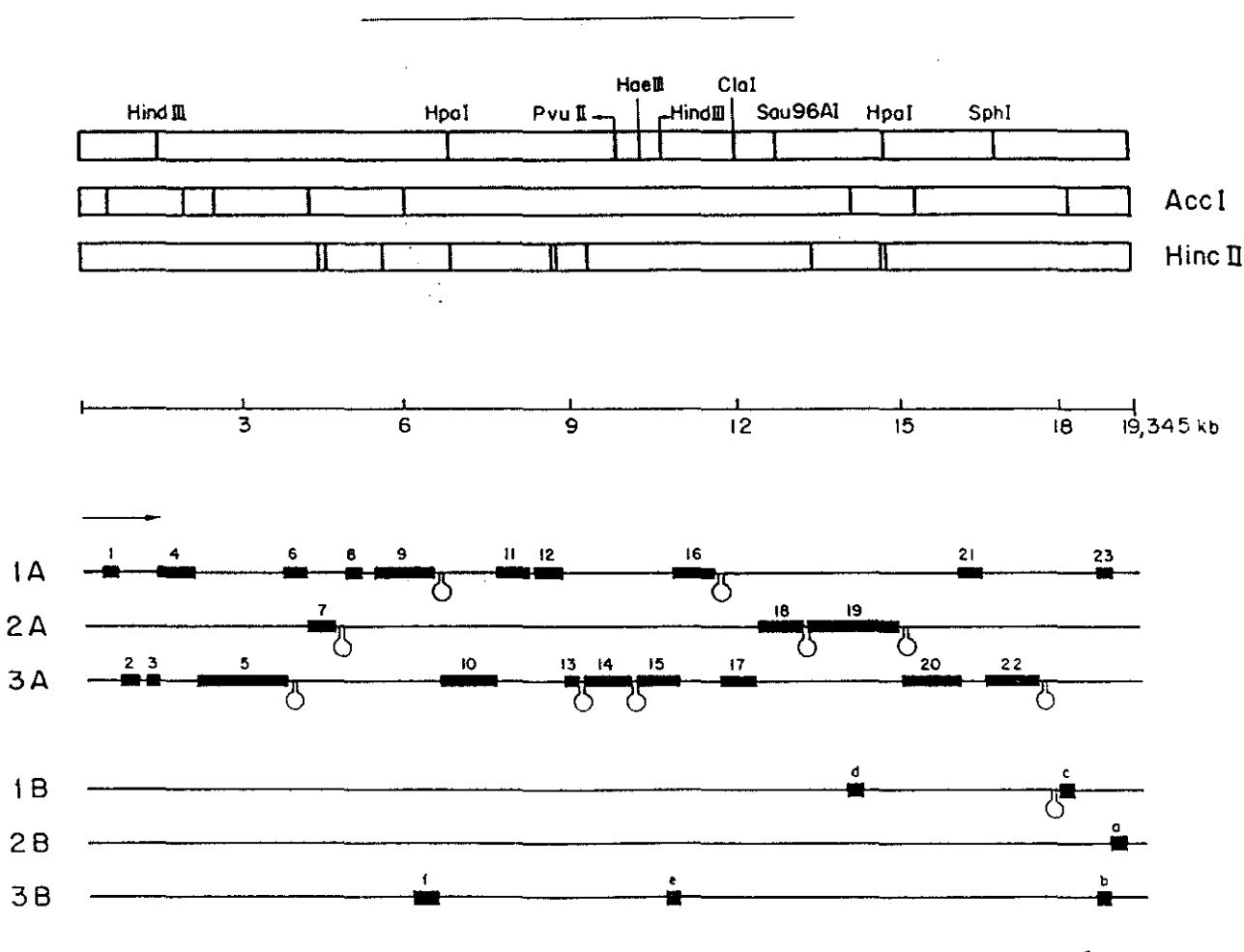


Figura 5. Mapa físico del DNA de Cp-1. En la parte superior se indican los sitios de corte de algunas de las enzimas de restricción. En la parte inferior se representan, en cajas negras y en las tres posibles fases de lectura, las ORFs derivadas de la secuencia de nucleótidos. \circ representan los presuntos terminadores de transcripción.

En la Tabla 7 se pueden ver algunas de las características de los productos de las ORFs descritas en el genoma del fago. Una peculiaridad que se puede observar en esta tabla es que la mayoría de las proteínas presentan un pI ácido. Las proteínas con pI básico son principalmente proteínas tempranas que, presuntamente están implicadas en replicación y algunas de ellas se unen al DNA. Estas características podrían explicar el carácter básico de tales proteínas.

Tabla 7. Características de la composición de aminoácidos de las proteínas codificadas por las ORFs de Cp-1

ORF	Nº de aminoácidos	pI	D + E ^a	R + K ^b	aminoácidos ácidos ^c (%)	aminoácidos básicos ^d (%)	aminoácidos polares ^e (%)	aminoácidos no polares ^f (%)
1	88	4,43	17	6	19,3	10,2	61,5	38,5
2	96	6,57	15	15	15,6	16,6	60,4	39,6
3	64	4,8	10	7	15,6	10,9	64,2	35,8
4	230	10,11	31	52	13,4	23,9	68,3	31,7
5	568	5,68	87	75	15,3	15,4	59,7	40,3
6	148	9,21	15	22	10,1	16,2	67,5	32,5
7	170	4,77	28	19	16,4	12,3	59,4	40,6
8	96	4,19	18	8	18,7	8,4	56,3	43,7
9	365	5,90	46	39	12,6	13,7	58,5	41,5
10	337	5,28	49	39	14,5	13,4	59,6	40,4
11	194	4,90	30	21	15,5	12,3	62,9	37,1
12	194	8,60	24	26	12,4	14,9	61,2	38,8
13	74	5,62	10	9	13,5	13,6	64,7	35,3
14	295	4,89	42	25	14,2	11,6	61,0	39,0
15	288	6,68	43	43	14,9	15,2	58,8	41,2
16	253	5,98	26	25	10,2	10,3	64,3	35,7
17	230	8,92	24	28	10,4	13,0	60,4	39,6
18	237	5,07	32	25	13,5	11,3	67,4	32,6
19	586	4,87	65	45	11,1	9,1	62,4	37,6
20	360	7,55	49	50	13,7	15,0	61,3	38,7
21	134	7,87	12	13	8,9	11,2	45,6	54,4
22	339	4,62	45	29	13,3	9,5	62,4	37,6
23	86	9,96	1	10	1,2	15,2	56,7	43,3
a	62	8,82	7	9	11,3	14,5	62,8	37,2
b	89	7,73	14	15	15,7	19,0	67,5	32,5
c	80	9,93	4	11	5,0	13,8	57,5	42,5
d	93	9,26	2	6	2,2	9,7	37,6	62,4
e	69	11,01	3	10	4,3	14,4	32,1	67,9
f	147	10,65	6	21	4,0	14,3	50,4	49,6

^a Aminoácidos con carga negativa (D, E) que contiene cada producto de la ORF.

^b Aminoácidos con carga positiva (R, K) que contiene cada producto de la ORF.

^c Porcentaje de aminoácidos ácidos (D, E) que contiene cada producto de la ORF.

^d Porcentaje de aminoácidos básicos (R, K, H) que contiene cada producto de la ORF.

^e Porcentaje de aminoácidos polares (R, K, H, D, E, N, Q, C, G, S, T, Y) que contiene cada producto de la ORF.

^f Porcentaje de aminoácidos no polares (A, V, L, I, M, P, W, F) que contiene cada producto de la ORF.

Por otra parte, en la Tabla 8 se indica la frecuencia de los codones que utiliza el fago para codificar los diferentes aminoácidos de las proteínas, producto de las 29 ORFs. A pesar de que no se conoce todavía la secuencia de un gran número de genes de neumococo, Martin y Claverys (1991) elaboraron una tabla de uso de codones para esta bacteria, basada en los datos obtenidos a partir de 28 genes estudiados. En general, no se observan grandes diferencias entre el uso de codones preferencial utilizado por neumococo y el utilizado por el fago. Hay tres casos en los que se pueden observar diferencias entre ambos en los codones que codifican la asparagina, la histidina y el aspártico. Neumococo utiliza preferentemente el codón AAT (67%) para codificar la asparagina mientras que el fago utiliza en similar proporción AAC (53,5%) que AAT (46,5%). Lo mismo sucede con los codones que codifican la histidina, el fago utiliza por igual CAT (50,1%) y CAC (48,9%) mientras que neumococo prefiere CAT (67%). Con los codones que codifican el aspártico, neumococo utiliza preferentemente GAT (66%) y el fago preferentemente usa GAC (61,5%). Estos datos parecen indicar que neumococo prefiere una T en la tercera posición del triplete del codón.

Tabla 8. Uso de codones de las 29 ORFs de Cp-1^a

K	AAG 177	H	CAT 45	D	GAA 274
	AAA 286		CAC 43		GAG 134
N	AAC 220	Q	CAA 138	F	TTT 251
	AAT 191		CAG 92		TTC 96
C	TGT 26	D	GAT 134	Y	TAT 191
	TGC 26		GAC 214		TAC 111
R	AGG 21	L	CTT 107	S	AGC 98
	AGA 66		CTC 31		AGT 58
	CGT 55		CTA 86		TCT 95
	CGC 48		CTG 43		TCC 21
	CGA 22		TTA 132		TCA 98
	CGG 28		TTG 101		TCG 14
V	GTT 112	G	GGT 74	A	GCT 132
	GTC 52		GGC 68		GCC 70
	GTA 117		GGA 90		GCA 116
	GTG 75		GGG 78		GCG 53
T	ACG 69	P	CCT 55	I	ATA 90
	ACA 128		CCC 25		ATC 144
	ACC 90		CCA 61		ATT 189
	ACT 76		CCG 25		
STOP	TAA 20				
	TGA 6				
	TAG 3				

^a Se indica el número de veces que aparece cada triplete. Los aminoácidos se indican en código de una letra.

1.3. Comparación de las secuencias de aminoácidos

La secuencia de aminoácidos codificados por cada ORF se comparó con las secuencias existentes en los bancos de datos SWISSPROT y PIR mediante el programa FASTA. El resultado obtenido fue que sólo se encontraron similitudes significativas con proteínas codificadas por el fago $\phi 29$ (Tabla 9). Dos de ellas están implicadas en los procesos de replicación, la TP y la DNA polimerasa, otras son proteínas estructurales y/o morfogenéticas (proteína mayoritaria de la cabeza, el conector, el collar, la cola y la proteína de encapsidación) y, por último, las implicadas en la lisis celular (la holina y la lisozima). Conviene citar, antes de continuar, que normalmente se habla indistintamente de conector y proteína portal por lo que a lo largo de esta Memoria se hablará de conectores. La similitud de estas proteínas de Cp-1 con las correspondientes de $\phi 29$ se mantiene a lo largo de toda la molécula con sólo dos excepciones. Una es el extremo C-terminal de las lisozimas, ya que como se ha comentado en la Introducción, la de Cp-1 presenta motivos de unión a colina que no están presentes en la de $\phi 29$. La otra excepción es la proteína de la cola. La proteína de la cola de $\phi 29$ presenta similitud con dos ORFs no consecutivas de Cp-1: el extremo N-terminal de la proteína gp9 de $\phi 29$ tiene un 73% de similitud con el producto de la ORF17 de Cp-1, mientras el extremo C-terminal de la misma proteína gp9 de $\phi 29$ es 72,6% similar a la ORF19 de Cp-1 (Figura 6). Curiosamente, la ORF18 presenta un 58,1% de similitud con una región conservada de la proteína 38 de varios bacteriófagos de *E. coli* como ox2, m1, t2 y k3 (Montag y cols., 1987). La proteína 38 está localizada en los extremos libres de las largas fibras de la cola y funciona a modo de adhesina, mediando la unión con los receptores de la bacteria. No se conoce si el producto de la ORF18 está implicado en un proceso similar, así como si las dos ORFs codificantes de las presuntas proteínas de la cola funcionan o no como péptidos separados.

La comparación de la secuencia de las TPs de varios sistemas que replican por el mecanismo de "protein-priming" muestra que su nivel de similitud es menor que el que presentan las TPs de $\phi 29$ y Cp-1. Sin embargo, los perfiles hidropáticos predichos para las TP de los diferentes virus ($\phi 29$, PRD1 y adenovirus) sugieren la existencia de una estructura secundaria similar con varias regiones hidrofílicas y muy pocas regiones hidrofóbicas (Shiue y cols., 1991). Se ha predicho la estructura secundaria para la TP de Cp-1 según el método de Chou-Fasman (Chou y Fasman, 1978) y se encontró que era similar a la postulada para

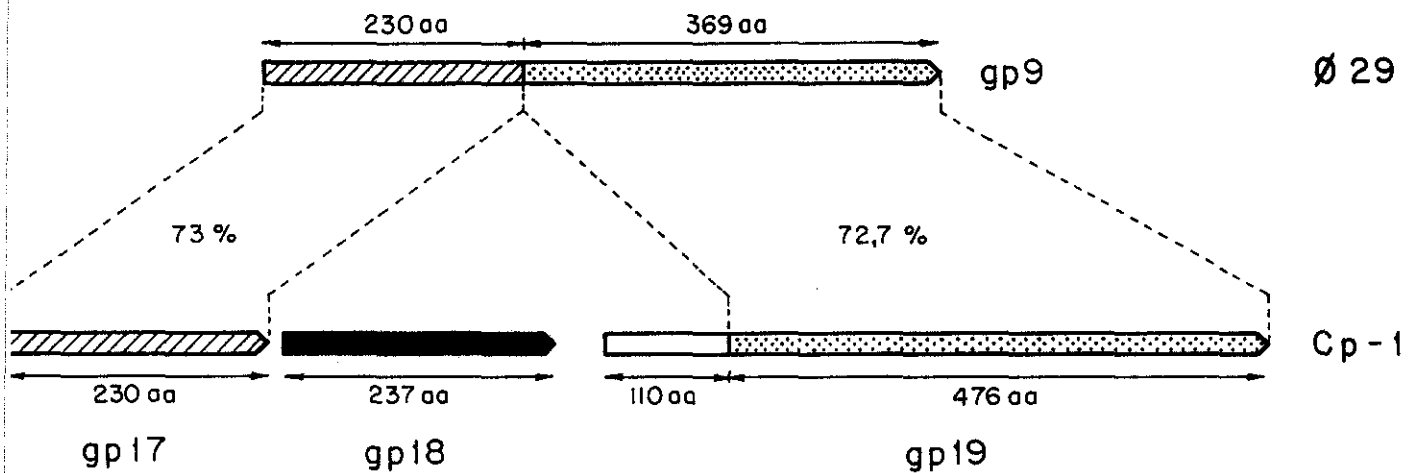


Figura 6. Esquema de las proteínas de la cola de $\phi 29$ y de Cp-1. Los diferentes rayados indican las regiones que presentan similitud entre la proteína de la cola de $\phi 29$ y los productos de las ORF17 y 19 de Cp-1. Se señala también el porcentaje de similitud entre dichas regiones. La flecha negra representa el producto de la ORF18 de Cp-1.

las otras TP's. Se ha demostrado, tanto en adenovirus (Freimuth y Ginsberg, 1986; Pettit y cols., 1989) como en $\phi 29$ (Zaballos y Salas, 1989), que la región N-terminal de las TP's está implicada en la unión al DNA y el C-terminal en la unión a la DNA polimerasa. El aminoácido de unión entre la TP y el DNA se ha caracterizado en varios genomas (Hermoso y cols., 1985; Shiue y cols., 1991; Smart y Stillman, 1982) pero en el caso de Cp-1 la treonina responsable de la unión continúa siendo materia de especulación.

Tabla 9. Comparación de proteínas de Cp-1 y de $\phi 29$

Proteína	aminoácidos idénticos (%)	aminoácidos conservativos (%)	Nº de aminoácidos		kDa	
			$\phi 29$	Cp-1	$\phi 29$	Cp-1
Proteína terminal	20,9	50,2	266	230	30,0	26,8
DNA polimerasa	37,1	58,6	572	568	66,5	66,2
Proteína de la cabeza	28,2	51,6	448	365	49,6	42,1
Conector	33,5	57,6	308	337	35,8	39,5
Collar	34,1	56,2	292	194	33,5	22,6
Cola ^a	24,6	48,4	599	230	67,6	26,2
Cola ^b	24,1	48,6	599	586	67,6	64,9
Encapsidación	32,6	54,9	332	360	38,8	42,2
Holina	20,2	50,0	134	134	14,8	15,4
Lisozima	20,7	43,8	257	339	26,8	39,3

^a Comparación entre la parte N-terminal del producto del gen 9 de $\phi 29$ y el producto del gen 17 de Cp-1.

^b Comparación entre la parte C-terminal del producto del gen 9 de $\phi 29$ y el producto del gen 19 de Cp-1.

El alineamiento múltiple de las DNA polimerasas secuenciadas hasta ahora ha permitido dividir las en tres grupos, A, B y C, basados en la homología con los productos de los genes *polA*, *polB* y *polC* que codifican para las DNA polimerasas I, II, y la subunidad α de la III de *E. coli*, respectivamente (Braithwaite e Ito, 1993). Atendiendo a esta

clasificación las DNA polimerasas que actúan en el mecanismo de "protein-priming" pertenecen al grupo B. Basado en estudios genéticos y bioquímicos se ha propuesto que las DNA polimerasas tienen, al menos, dos dominios separados, uno es el dominio 3'→5' exonucleasa y el otro el dominio de polimeración, situados en el extremo N- y C- terminal respectivamente (Bernard y cols., 1987; Joyce, 1991). La DNA polimerasa postulada para el Cp-1 presenta, como era de esperar, una gran similitud con las enzimas pertenecientes al grupo de genomas que replican por el mecanismo de "protein-priming". En particular, presenta un 37% de aminoácidos idénticos con la DNA polimerasa de $\phi 29$, y este valor aumenta a 95,7%, cuando se consideran también los aminoácidos conservativos. En la Figura 7 se puede ver una comparación de los distintos motivos conservados que se encuentran en ambos dominios. En dicha comparación se incluyen diferentes enzimas de este grupo tanto de eucariotas como de procariotas, así como la DNA polimerasa II de *E. coli*, la cual es estructuralmente homóloga a las DNA polimerasas de tipo α (Ishino y cols., 1994). Además, existen dos regiones altamente conservadas entre las regiones 1 y 2a, un aspecto típico de las DNA polimerasas de los genomas que replican por "protein-priming". Se ha postulado que estas regiones puedan estar relacionadas con este mecanismo de replicación. Asimismo, hay que resaltar que la similitud encontrada entre las DNA polimerasas del fago Cp-1 y de su huésped es mucho menor (19% de aminoácidos idénticos), que las ya señaladas entre sistemas que replican por "protein-priming". El hecho de que se parezcan más las DNA polimerasas de los fagos Cp-1 y $\phi 29$ que la del fago con la de su propio huésped no es sorprendente puesto que ambos fagos comparten un mismo mecanismo de replicación que es distinto al de la bacteria.

El producto de la ORF20, como se muestra en la Tabla 9, presenta similitud con la proteína de la encapsidación de $\phi 29$. Se ha demostrado que esta proteína es una ATPasa (Guo y cols., 1987d) que proporciona la energía necesaria para el empaquetamiento del DNA del fago en la precabeza durante el proceso de morfogénesis, mediante la hidrólisis de ATP. La ATPasa se une a la precabeza a través de una molécula pequeña de RNA (pRNA) unida al conector de las precabezas (Guo y cols., 1987c; Donate y cols., 1993), mediante un dominio de unión a RNA (Donate y cols., 1993), sufriendo cambios conformacionales que le permiten unirse al complejo TP-DNA y al ATP. Después, se produce la hidrólisis del ATP y la consiguiente traslocación del complejo TP-DNA a la precabeza. Esta ATPasa contiene las dos secuencias consenso de unión a ATP, la de tipo A y la de tipo B (Figura 8).

		EXO I		EXO II		EXO III
PolII	153	SI I I RHG	220	IGW VVQ LRM	327	ALAT NLK CE
Adeno	138	TY V I WMG	271	VG ING E V	434	T CAL VQ
pGKL1	361	CF I SYFDP	427	LIA GGGY FHYI	559	WSK CRN L
S1	210	FVA L LLLKR	294	V SO MI	393	DS T L Q IL
PRD1	14	AAF DPFKH	67	A GG F	141	I E L G CV
φ29	6	YSC KV	58	L A I	161	EYA I N IQ
Cp-1	4	YAG NE	68	F S	135	WI IHV A

		REGION 1		REGION 2a		REGION 2b
PolII	417	L YK S I I R T F L I D V	490	GAL M AF VLG	523	GHQIM QT
Adeno	541	ICGM A TH-PM W	692	SI LLS A S	835	LA T AF
pGKL1	647	CL K A FY-DQ Y	781	NVI M FW	848	TLS YH
S1	483	Y S DD--M I	614	FIY TM R G	696	GY IR
PRD1	218	M HA R NFR-H FS	337	IFY L S	406	I GA SV
φ29	247	F Q Y SRL-L Y	380	L LM	433	I A YT
Cp-1	245	I I M T QNA-L I	383	K M G	436	SI HF

		REGION 3		REGION 4
PolII	539	YDV G T	595	TH CRFLMPTI GADTGSK R
Adeno	864	LKSV G L	924	DAYSPE--VFAP-----
pGKL1	868	AEC S	911	CE D L--LCAGK-----Q
S1	711	DCY T VV	743	HFVEYG--IFAP-----S
PRD1	422	ERPL C IC	453	ATGD I--AIAGK-----
φ29	450	RI C HL	486	ST KRA--KY Q-----T
Cp-1	453	NFL A LHL	486	GRAV A--KY S-----

Figura 7. Regiones conservadas de las DNA polimerasas en los dominios 3'→5' exonucleasa y de polimerización. El alineamiento de las secuencias de aminoácidos se obtuvo usando el programa PILEUP del GCG (Genetics Computer Groups). Las secuencias de las DNA polimerasas que aparecen junto con la del fago Cp-1 se obtuvieron de Braithwaite e Ito (1993). Los números indican la posición con respecto al extremo N-terminal de cada DNA polimerasa. Las cajas negras representan los residuos idénticos en al menos cinco de las siete secuencias comparadas. Otras similitudes se indican en las cajas sombreadas. Las regiones EXO I-III pertenecen al dominio 3'→5' exonucleasa y las regiones 1-4 al de polimerización.

Estos motivos están presentes en otro tipo de ATPasas y proteínas que unen ATP como la ATPasa α y β de *E. coli* o la adenilato kinasa, aunque no todas presentan ambos motivos, sino sólo uno de ellos (Guo y cols., 1987d). Las proteínas encargadas de empaquetar el DNA en otros fagos como lambda, T7 o T4 presentan también dominios de unión a ATP, pero sólo uno de ellos (Guo y cols., 1987d). El producto de la ORF20 también muestra ambos motivos de unión a ATP (Figura 8) por lo que posiblemente se trata de la ATPasa encargada de empaquetar el complejo TP-DNA de Cp-1 en la precabeza fágica. Además, se puede sugerir la presencia de un dominio de unión a RNA en la proteína codificada por la ORF20 de Cp-1, dada la alta similitud que presentan las secuencias de ambas proteínas.

Por otra parte, en la Tabla 9 no se ha incluido la ORF8 que codifica una proteína de 10,5 kDa y que presenta una identidad del 20% y una similitud del 46% con la proteína de andamiaje de ϕ 29. Estas proteínas, sobre las que se forman las cápsidas de los fagos, no presentan similitud a nivel de secuencia de aminoácidos entre los diferentes sistemas aunque a veces sí a nivel de estructura terciaria (por ejemplo la de los fagos T7, T4, P22 y lambda presentan en su secuencia N-terminal un gran contenido en hélice α) (Casjens y Hendrix, 1988). Este hecho, junto con el orden funcional de los genes implicados en el ensamblaje de la cabeza, donde la proteína de andamiaje se encuentra precediendo a la proteína mayoritaria de la cabeza en los genomas de fagos de dsDNA estudiados hasta ahora (Eppler y cols., 1991; Hatfull y Sarkis, 1993; Duda y cols., 1995a, b), nos lleva a sugerir que es muy probable que la ORF8 codifique la proteína de andamiaje del Cp-1. No obstante, el grado de similitud de esta proteína con la del ϕ 29 no es tan alto como el de otras proteínas estructurales y morfogenéticas y, en este caso, la predicción de sus perfiles hidropáticos difiere sustancialmente, aunque ambas proteínas tienen un gran contenido en hélice α en su estructura secundaria, 49 % para la de ϕ 29 y 57,3 % para la de Cp-1, siendo el resto sustancialmente, estructura al azar (40,8 % en la de ϕ 29 y 31,2 % en la de Cp-1). Además, casi todas las proteínas de andamiaje conocidas están muy cargadas al igual que esta propuesta para el Cp-1 (27,1% de aminoácidos cargados), por ejemplo las de T4 presentan un 45,9 y 33,3% de aminoácidos cargados, la de T7 30,3%, la de ϕ 29 40,2% (Eppler y cols., 1991), las cuales también presentan un alto contenido en estructura en hélice α .

A

Proteína	Residuos	Secuencia									
gp 19 de T7	51-72	K	KF-ILQ	A	FR	G	I	GKS	FITCAFV	V	
gp 17 de T4	154-174	R	MTVCNL	S	-R	Q	L	GKT	T-VVAIF	L	
	293-314	H	FY-DIW	T	AA	V	E	GKS	GFEPYTA	I	
gp A de lambda	484-505	R	VI-PIK	G	AS	V	Y	GKP	VASMPRK	R	
gp 9 de PRD1	11-30	R	IL--VL	G	HT	G	T	GKT	CAA-VWH	L	
gp 16 de ø29	17-39	R	ILNFEVI	G	AR	G	I	GKS	YAMKVYP	I	
gp 20 de Cp-1	18-39	Q	YLNFEVI	G	GR	G	Y	GKT	-ALKKYL	F	

B

Proteína	Residuos	Secuencia									
Fosfofructokinasa	90-102	K	-----K-H	G	IQG	L	VVIGG	D			
ATPasa β de <i>E. coli</i>	232-243	R	-----D-E	G	RDV	L	--LFV	D			
ATP/ADP translocasa	279-291	R	-----GMG	G	AFV	L	-VLY-	D			
Adenilato kinasa	107-119	R	-----K-I	G	QPT	L	-LLYV	D			
ATPasa β bovina	246-258	R	-----DQE	G	QDV	L	--LFI	D			
gp 16 de ø29	248-256	K	FVFSIVYN	G	-FT	L	-GVWV	D			
gp 20 de Cp-1	283-292	R	-LYSFYCD	G	-TF	L	TGYII	D			

Figura 8. Secuencias consenso de los dominios de unión a ATP, tipo A (A) y tipo B (B). Las cajas blancas representan aminoácidos básicos y las negras aminoácidos o regiones hidrofóbicas. Se recuadran los aminoácidos conservados en los dominios. Las secuencias que aparecen están tomadas de Guo y cols. (1987d).

2. MAPA DE TRANSCRIPCIÓN DEL BACTERIÓFAGO Cp-1

Una vez obtenida la secuencia completa del genoma nos propusimos elaborar un primer mapa de transcripción del fago. En este punto, hay que recordar que la temperatura óptima de replicación del fago es 30°C y la adsorción fágica es muy lenta de manera que, después de 10 min de iniciada la infección, sólo se han adsorbido un 20% de los fagos y, hasta después de una hora, no se adsorben el 50% de los mismos (Ronda y cols., 1981). Por tanto, para realizar los ensayos de transcripción se obtuvo RNA de neumococo infectado con Cp-1 a diferentes tiempos después de la infección, tal y como se describe en el apartado 13 de Materiales y Métodos. También se obtuvo RNA de cultivos de neumococo sin infectar como control de especificidad de los transcritos de Cp-1. Para determinar el número y el tamaño de los mRNA, en primer lugar se hibridaron muestras de estos RNAs con el genoma completo mediante la técnica del "Northern-blot". Posteriormente, para localizar estos transcritos en el genoma, se realizaron hibridaciones con sondas específicas de diferentes regiones del genoma. Por otra parte, para determinar el sitio exacto de iniciación de la transcripción mediante la técnica de "primer extension". Como resumen de estos experimentos se pudo concluir que los primeros RNAs transcritos denominados tempranos no aparecieron hasta los 20 min, aproximadamente, después de la infección, aumentando su expresión hasta alcanzar un máximo alrededor de los 40 min y después los niveles de transcripción, fueron descendiendo lentamente hasta los últimos estadios de la infección. La otra categoría de RNAs transcritos (tardíos) se empezaron a detectar aproximadamente a los 35-40 min después de la infección y alcanzaron su máximo nivel de expresión a los 60 min. Al igual que los primeros, su expresión fue disminuyendo lentamente y se siguieron detectando hasta que tiene lugar la lisis. La larga vida media de estos mRNAs fágicos dió lugar a un abundante rastro en muchos de los "Northern-blot" analizados, lo que dificulta en ocasiones la asignación precisa de algunos de los mismos (Martín y cols., 1995, 1996). El rastro que se observa en estos experimentos puede ser el resultado del procesamiento no específico de los transcritos fágicos, o la degradación de los transcritos, o la presencia de transcritos con un extremo 5' común pero que tienen tamaños diferentes o una combinación de todos los anteriores. Se hicieron también ensayos de "Northen-blot" tratando los cultivos con Cm previamente a la infección, para determinar los mRNAs tempranos.

2.1. Determinación del sitio de iniciación de la transcripción

Para poder situar con precisión el punto de iniciación de los diferentes RNAs transcritos de los genes fágicos se hicieron ensayos por "primer extension". Para ello se analizó la secuencia del genoma fágico confirmando la existencia de una serie de regiones con secuencias presuntamente promotoras, en las que se incluyen no sólo las secuencias canónicas que presentan las regiones -10 y -35 claramente identificables, sino también las secuencias que contienen la región -35 pero carecen de la caja -10, y viceversa. Para comprobar si estas secuencias son funcionales se hicieron los experimentos de "primer extension", mediante la utilización de oligonucleótidos con una secuencia complementaria a la región situada a unos 100 pb del extremo 5' del presunto transcrito. La secuencia de los oligonucleótidos y su localización en el genoma se resume en la Tabla 10.

Tabla 10. Oligonucleótidos utilizados en los ensayos positivos de "primer extension"

Oligonucleótidos	Secuencia 5'→3'	Localización ^a	ORF ^b
PEX150	AACATCTAGCGCTTGCCCGTG	453-433	1
PEX151	TTCATGTTTCATGCCTGCTTGC	18.919-18.939	a
PEX152	CAGCGTCTAGCACTGCGAGGC	4.912-4.892	8
PEX153	GTAAGGAAGCTGGCGACGTTGG	1.225-1.184	3
PEX154	AAGAATTGTGATATAGTCGCC	16.080-16.060	21
PEX155	TTACCAATTCCACGACCACCG	15.046-15.026	20
PEX156	GGCTAGACGGTTACGCTCTACGG	6.641-6.619	10
PEX158	CCCCCGATGTCGCAACCGCCC	13.212-13.192	19
PEX161	CATAATCGTTTTTAAAGGGC	11.646-11.626	17
PEX162	TCTTGAAAGCTGAAAAGCGTG	8.300-8.280	12
PEX170	ATTCTTGAGGGGGTAACTCCTC	18.669-18.648	23

^a Coordenadas de los oligonucleótidos en el genoma del fago.

^b ORF en la que se localiza el oligonucleótido.

Inicialmente se analizaron los transcritos que se podían originar a partir de las ORFs situadas en ambos extremos del genoma. Así, el análisis realizado con el oligonucleótido PEX150 (complementario a la ORF1) mostró dos bandas nítidas en las posiciones 257 y 351, que no estaban presentes cuando se usó el RNA extraído de cultivos de neumococo sin

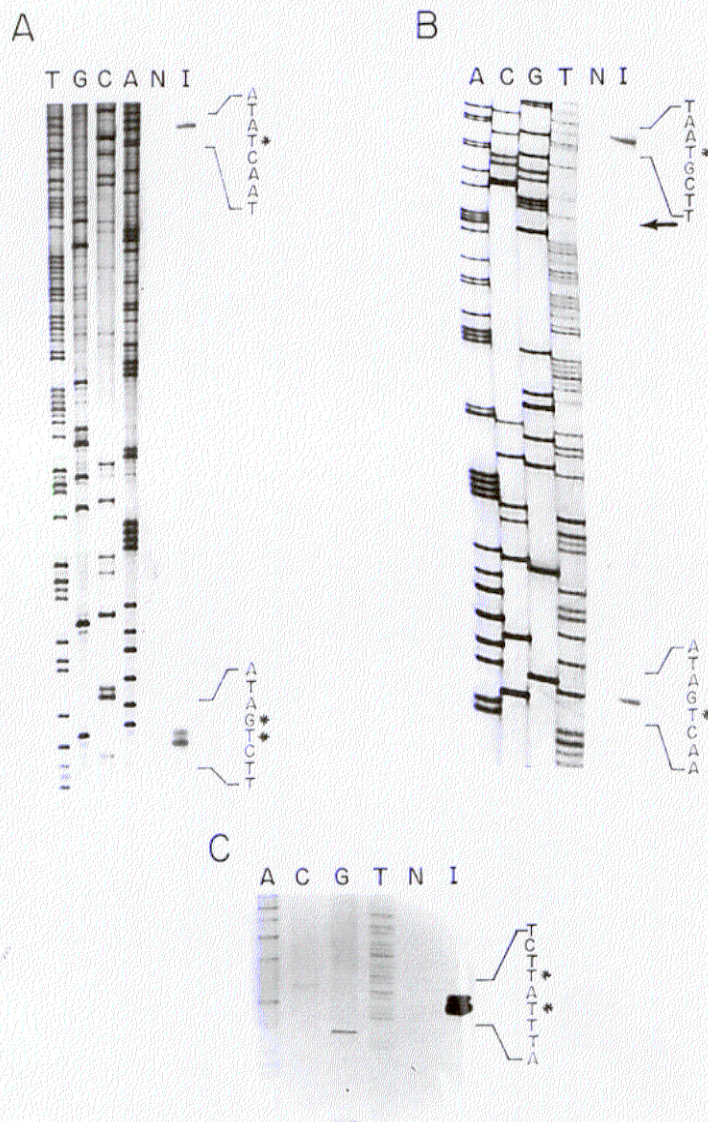
infectar (Figura 9B). La secuencia control fue, en este caso, la de un plásmido recombinante que contenía un fragmento del genoma del fago hasta el nt 278 y después se continuaba con la secuencia múltiple de clonación del vector. No obstante, comparando la secuencia del vector con la del genoma del fago se puede asignar la banda de mayor peso molecular, que corresponde al nt 257. Precediendo estas bandas se detectan dos regiones promotoras consenso de *E. coli* (-35: TTGACA, 16-19 nt de separación, -10: TATAAT, 5-9 nt hasta el sitio de iniciación de la transcripción) (Harley y Reynolds, 1987) y es interesante resaltar que la primera secuencia promotora y la caja -35 de la segunda se encuentran dentro de las ITRs. Estos dos sitios de iniciación de la transcripción están separados 94 nt y sus intensidades fueron similares, lo que indica que la RNA polimerasa es capaz de unirse y de iniciar la síntesis de mRNA en ambos sitios aproximadamente con la misma eficacia.

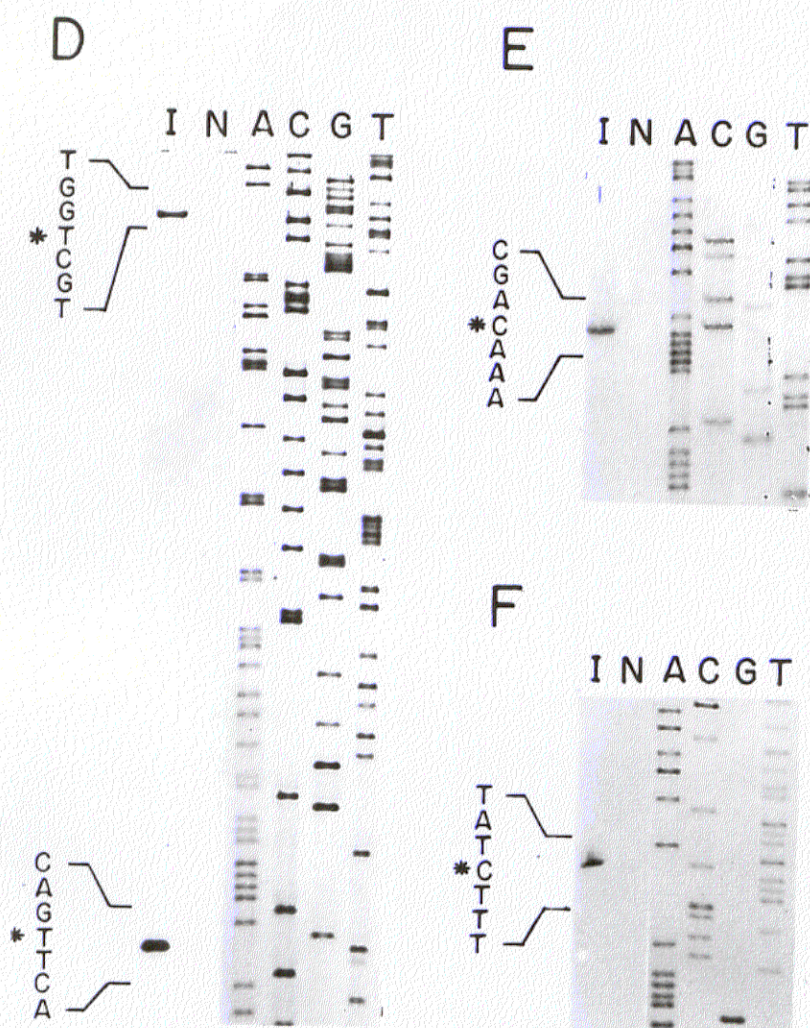
En la Figura 9C, se puede observar el análisis realizado a partir del híbrido RNA-PEX153 (complementario a la ORF3), observándose una banda en el nt 1.005 y una segunda banda en el nt 1.007. En este caso, la molécula más grande será la que probablemente se produzca en la transcripción *in vivo*. La secuencia promotora que precede a estos nt de iniciación no concuerda totalmente con las secuencias de las cajas -35 y -10 de las secuencias consenso de los promotores descritos para *E. coli* (Tabla 11).

El análisis realizado a partir del PEX151 (complementario a la ORFa) hibridado con el RNA total, dió como resultado la detección de tres, dos de las cuales forman un doblete, bandas que se corresponden con los nt 18.995, 18.996 y 19.089, también separados 94 nt (Figura 9A) y cuyas secuencias promotoras precedentes están perfectamente conservadas con respecto a las consenso de *E. coli* (Tabla 11). Al igual que en el caso del oligonucleótido PEX150, la primera región promotora y la caja -35 de la segunda se encuentran localizadas en las ITRs, como era de esperar dada su localización en las ITRs.

Por otra parte, usando como cebador el PEX152 (complementario a la ORF8) se obtuvieron dos bandas de similar intensidad en los nt 4.740 y 4.835, separadas por 95 nt, por lo que suponemos que los dos sitios de iniciación existen *in vivo*, como ya se ha citado antes para el caso de los promotores en tándem en los extremos del genoma (Figura 9D). Precediendo a ambos sitios de iniciación existen secuencias promotoras similares, aunque no totalmente conservadas respecto a las canónicas. Asimismo, con el PEX154 (complementario a la ORF21) se observaron dos bandas de similar intensidad, que corresponden a los nt 15.908 y 15.984 (Figura 9G). Con los oligonucleótidos PEX156,

PEX162, PEX158 y PEX155 (complementarios a las ORFs 10, 12, 19 y 20, respectivamente) se obtuvieron bandas únicas de transcripción en los nt 6.535, 8.193, 13.111 y 14.938, respectivamente (Figura 9I, E, H, F). Con el oligonucleótido PEX161 aparece un doblete en los nt 11.570 y 11.571 (Figura 9J). Estos resultados demostraron que precediendo a las ORF8, 10, 12, 17, 19, 20 y 21 existe un sitio de iniciación de la transcripción y, por tanto, el comienzo de, al menos, un mRNA.





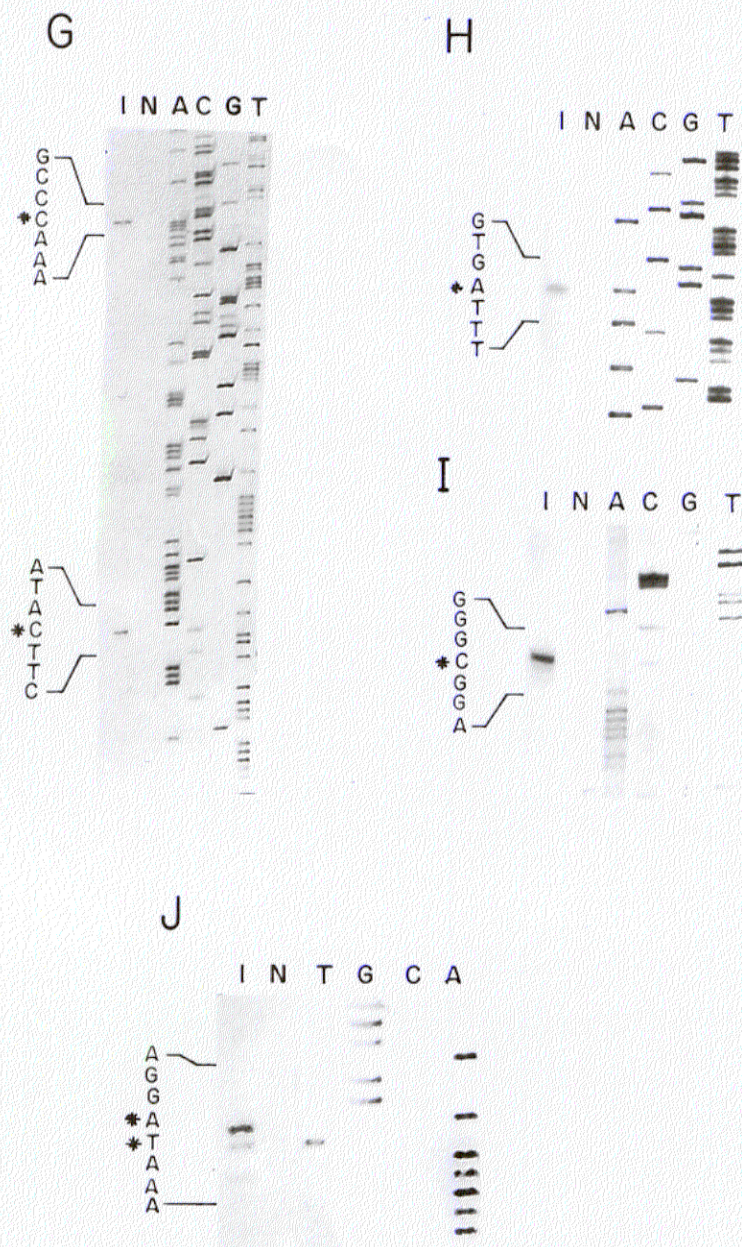


Figura 9. Determinación de los sitios de iniciación de la transcripción. Los ensayos de “primer extension” se realizaron usando RNAs de células sin infectar (calles N) o infectadas (calles I). Los oligonucleótidos que aparecen en la **Tabla 10** se anillaron con el RNA y se extendieron con la transcriptasa inversa para la identificación de distintos promotores: **A)** PE4-PE5. **B)** PE1-PE2. **C)** PE3. **D)** PL1-PL2. **E)** PL4. **F)** PL7. **G)** PL8-PL9. **H)** PL6. **I)** PL3. **J)** PL5. Una reacción de secuencia (ACGT) se cargó en el gel, en paralelo con la reacción del “primer extension”. Los asteriscos representan los sitios de iniciación de la transcripción. La flecha en A representa el sitio donde comienza la secuencia de pUC18.

En la Tabla 11 se resumen todas las secuencias promotoras funcionales existentes en el genoma del fago así como su localización. En ella, se han incluido también las secuencias consenso de las regiones promotoras de los fagos de *B. subtilis* (Murray y Rabinowitz, 1982) y de los genes conocidos de neumococo (Morrison y Jaurin, 1990) que muestran una secuencia similar a la de los promotores de *E. coli*. Como se aprecia en dicha tabla, se han detectado 15 promotores funcionales, ocho de los cuales forman cuatro pares de promotores en tándem. La distancia entre los dos sitios de iniciación de transcripción de dichos tándem varía entre 76 y 95 nt. El denominado PPR es el responsable de la síntesis de un RNA corto y no traducido del que se hablará más adelante. La mayoría de los 15 promotores muestran las típicas cajas consenso -10 y -35 reconocidas por la RNA polimerasa del huésped, aunque hay algunos que carecen de alguna de estas cajas. En el análisis global de todos los promotores se puede observar que los que corresponden a genes tempranos, que veremos a continuación, muestran un mayor grado de similitud entre las cajas y las regiones existentes entre ellas, con lo que se puede postular una secuencia consenso, mientras que los promotores de los genes tardíos muestran una mayor heterogeneidad. Otro dato interesante es la presencia, en ocho de los 15 promotores, del dinucleótido TG en la región -10: TGNTATAAT (Tabla 11), lo que se ha denominado "región extendida -10". Este dinucleótido TG asociado a la región -10 del promotor es perfectamente reconocido por el factor sigma 70 en *E. coli* aunque en estos promotores no esté presente la región -35 (Kumar y cols., 1993). Recientemente se ha postulado que los promotores de *S. pneumoniae* requieren con frecuencia la presencia de esta extensión de la caja -10, pudiendo, en tales casos, ser funcionales sin la región -35 (Sabelnikov y cols., 1995).

Tabla 11. Promotores y sitios de iniciación de transcripción de Cp-1

Promotor	-35	-10	Espacio ^a	Localización ^b
Fagos de <i>Bacillus</i> ^c	TTATNAAAAAGTG TTGACA	ATNNTTACAAGGTGTGG TATAAT ANAAT	17	---
<i>S. pneumoniae</i> ^d	NNTCNGAAANNAC TTGACA	AANNCNNANANTNNNGC TATAAT NNNGN	17	---
PE1	TTTCAAAAATCTC TTGACA	ATCATTTT TAGTAAATGC TATAAT GGTATAGTTAGAA	17	ORF1
PE2	TTTCAAAAACAC TTGACA	AGCATTTT GAGAAAATGC TATAAT AGTATCAGTTAAAA	17	ORF1
PE3	TTAAGTTTATCAC TTGCTT	TCTATTTT GAAAATGTGG TAAACT TAAATAAGAAAAAT	17	ORF3
PE4	TTTCAAAAATCTC TTGACA	ATCATTTT TATAAAATGA TACAAT GTATATAGTTAGAA	17	ORFa
PE5	TTTCTTTTTTCTC TTGACA	AGCATTTT GAGAAAATGC TATAAT GGTATCAGAAAGAT	17	ORFa
PL1	TTATAAAGCGTTT TTTCATG	GCTTGAAAAGTCAGTTGGT TAGAAT GACTCGCACCAGCA	19	ORF8
PL2	TTGCCGAAAAGCG TTACAG	GTTTATCATATCATACT TGC TTTATT TTGTCAAGTATGAT	20	ORF8
PL3	CATTTCTGGTATG TTGAGG	AAGCCTAAAGGCTGACCAAGGGCG GGCAAT AGCCCGCCTATTTT	24	ORF10
PL4	AAGGATATACAGT TTCAGA	TATTTAATATATGT GATAAG CTGTTTTTACAAGT	14	ORF12
PL5	GACAAGGACAGAT AGGACA	AGAATAAGGGAGGG TTTCAG CCTCCTATTTT	14	ORF17
PL6	GAAAGGGTAAACA ATGGCA	GAAGCAACAGAACGC TAAAAG CACTAAATGAAATC	15	ORF19
PL7	TTTGATATTCAGA AAAACA	ACCCGAATTTTAAAAAGTATGA TATAAT GAAATAGAAAGGAG	22	ORF20
PL8	CTTTGACAAAAAT TTGACC	GCTTATTGTTTAGAAG TGGAAC AGGTGGAAGACGGG	16	ORF21
PL9	CACTAGGTAAAC TTTACG	GAGTTTGTACTTAAAAAA TATGTT TATTTATGAAGATT	18	ORF21
PPR	ATAATAAGCGTGT TTAAGC	GTCATTTT TAAAAATAGT TCTCAT GGACTTCTTTATAT	18	dentro de la ORF23

Los nucleótidos que coinciden con los de las cajas -35 y -10 consenso de *E. coli* y con la región extendida de la caja -10 aparecen en negrita. $\square \rightarrow \square$ indican los sitios de inicio de transcripción.

^a Distancia, en nt, entre las cajas -35 y -10.

^b Localización de los promotores respecto a las diferentes ORFs que preceden.

^c Secuencia consenso de los promotores tempranos de los genes de los fagos de *B. subtilis* (Murray y Rabinowitz, 1982).

^d Secuencia consenso de los promotores de los genes de neumococo (Morrison y Jaurin, 1990).

2.2. Mapa de transcripción de los genes tempranos

La DNA polimerasa y la TP de Cp-1 juegan un papel fundamental en el peculiar mecanismo de replicación de su DNA. La localización de los genes que codifican estas proteínas en la parte izquierda del genoma concuerda con la organización genética de otros sistemas fágicos (Salas, 1991). La expresión de tales genes junto con aquellos otros que intervienen en la replicación se realiza en las primeras etapas de la infección, lo que requiere una expresión temprana (Murray y Rabinowitz, 1982). Para confirmar la validez de este hecho en el caso de Cp-1 se hicieron sondas que abarcaran distintas zonas de esta región, como se muestra en la Tabla 12, con el fin de hibridarlas con muestras de RNA obtenidas a los 20, 40, 60, 80, 100 y 120 min después de la infección. Con la sonda 1, que comprende la parte del genoma que contiene las ORFs 4 y 5, se obtuvieron dos bandas principales de hibridación de 4,5 y 3,6 kb, y una más débil de 2,6 kb (Figura 10). Como se puede observar en la Figura 10 las intensidades de las bandas varían en los diferentes tiempos ensayados indicando así la cinética de aparición de estos RNAs transcritos. Estos transcritos empiezan a aparecer a los 20 min, y, como se ha citado anteriormente, su expresión va aumentando hasta alcanzar el máximo alrededor de los 40 min, para disminuir lentamente hasta las etapas finales de la infección.

Para asignar estas bandas de hibridación a las ORFs del genoma se prepararon otras sondas específicas de esta región, y se analizó la secuencia para encontrar posibles regiones promotoras y/o terminadoras. La hibridación de los RNAs con la sonda 2, que comprende parte de las ORFs 1 y 2 mostró una sola banda de hibridación de 4,5 kb, mientras que con la sonda 3, que comprende parte de las ORFs 6 y 7, pudieron observarse dos bandas de 4,5 y 3,6 kb (Figura 10). Para determinar con más precisión el final de estos transcritos se hicieron "Northern blot" con una sonda específica de la región inmediatamente adyacente al extremo 3' de la ORF7, y se comprobó que no aparecían estas bandas. Existen dos posibles terminadores de transcripción en esta región, situados en los extremos 3' de las ORFs 5 y 7, con un ΔG de -3,4 y de -14,5 kcal/mol, respectivamente (Figura 5). A la vista de estos resultados de análisis por "Northern-blot" y "primer extension" se puede concluir que:

a) la banda de 4,5 kb corresponde a un mRNA que comienza en la zona que precede a la ORF1 y que acaba en el extremo 3' de la ORF7.

b) la banda de 3,6 kb determina un mRNA desde el comienzo de la ORF3 hasta el final de la ORF7.

c) la banda de 2,6 kb, podría corresponder a un mRNA más pequeño que comprendería a los genes de la TP y de la DNA polimerasa, es decir las ORFs 4 y 5, respectivamente (Figura 10).

Tabla 12. Sondas utilizadas para los ensayos de "Northern-blot"

Sonda	Localización ^a
pGG2	9.907-12.821
pGG4	10.733-13.438
pGG6	6.864-10.736
pGG9	5.393-6.554
pGG11	11.437-13.013
pGG12	9.807-12.874
pGG14	9.961-12.195
pGG15	11.851-12.199
pGG18	9.553-10.964
pGG37	5.512-8.551
pGG38	6.402-7.905
pGG40	6.628-7.688
pGG41	5.522-7.545
pGG52	5.037-6.016
pGG66	613-1.021
pGG76	13.291-15.950
pGG77	18.074-18.954
pCIP50	15.353-18.215
sonda 1 ^{b, c}	1.338-3.845
sonda 2 ^{b, c}	489-941
sonda 3 ^{b, c}	3.834-4.449
sonda I (pGG61) ^d	276-1.452
sonda II ^{b, d}	5.500-7.780
sonda III ^{b, d}	9.980-10.964
sonda IV (pGG74) ^d	13.048-14.916
sonda V ^{b, d}	15.985-17.556

^a Coordenadas del fragmento en el genoma de Cp-1.

^b El fragmento utilizado como sonda se obtuvo mediante la amplificación del DNA de Cp-1 por la técnica de PCR.

^c La sonda indicada se utilizó en los experimentos de "Northern-blot" que aparecen en la **Figura 10**.

^d La sonda indicada se utilizó en los experimentos de "Northern-blot" que aparecen en la **Figura 11A**.

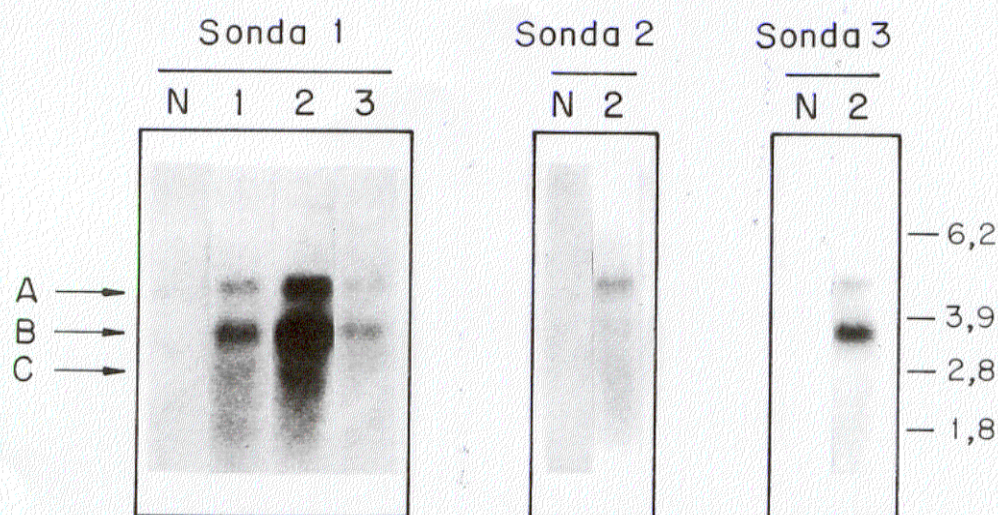


Figura 10. “Northern-blot” del RNA de células de neumococo infectadas con Cp-1. El RNA se obtuvo de células sin infectar (calles N), y a distintos tiempos después de la infección: 20 min (calle 1), 40 min (calles 2), o 60 min (calle 3). Las membranas se hibridaron con las sondas indicadas que aparecen en la Tabla 12. Se cargaron aproximadamente 10 μ g de RNA en los geles que se hibridaron con la sonda 1, y alrededor de 5 μ g para la 2 y la 3. Las flechas indican los diferentes mRNAs detectados y las letras corresponden a los mRNAs que aparecen en la **Figura 11B**. Los marcadores de peso molecular, en kb, se muestran a la derecha.

Estos resultados indican que las ORFs 3, 4 y 5 se transcriben juntas por medio de dos diferentes mRNAs tempranos (mRNAs A y B) (Figura 10), siendo predominante el que se transcribe a partir del promotor PE3, que comienza en el extremo 5' de la ORF3 y termina en el extremo 3' de la ORF7. Por otro lado, no está claro si la débil banda de 2,6 kb que va desde la ORF3 hasta el posible terminador en el extremo 3' de la ORF5 (mRNA C),

representa un verdadero mRNA o es el resultado de la degradación del mRNA predominante de 3,6 kb. Las ORFs 1 a 7 se transcriben juntas con un único mRNA (mRNA A) (Figura 10) que comienza en el extremo 5' de la ORF1 y termina en el extremo 3' de la ORF7 a partir de los promotores en tándem PE1/PE2.

Por otra parte, cuando se utilizó como sonda de hibridación de los RNAs el plásmido pGG61 (sonda I) (Tabla 12) que contiene 70 pb pertenecientes a las zonas de homología entre ambos extremos, se obtuvieron las mismas bandas anteriores más una banda adicional de 1,3 kb (Figura 11A). Este nuevo mRNA sería el que corresponde a las ORFs a, b, c y la dirección de transcripción sería de derecha a izquierda (mRNA D) (Figura 11A). La cinética de aparición de este mRNA fue la misma que la de los anteriores, por lo que se trata de otro mRNA temprano. Hay que recordar que precediendo a la ORFa se encuentran dos regiones promotoras en tándem funcionales (PE4/PE5), como se ha comprobado en el apartado anterior cuando se usó el oligonucleótido PEX151.

Por tanto podemos concluir que entre los genes tempranos se encuentran los que codifican, entre otros, la DNA polimerasa y la TP, proteínas implicadas en la replicación del DNA. El producto de los otros genes no se conoce todavía pero atendiendo a su localización con respecto a otros genomas fágicos como $\phi 29$ o PRD1 y por su expresión temporal podría suponerse que codificarían proteínas implicadas en procesos de replicación, tales como las proteínas de unión a DNA de simple y doble banda, y/o proteínas implicadas en el control de la transcripción tardía (Salas y Rojo, 1993; Bamford y cols., 1995).

2.3. Mapa de transcripción de los genes tardíos

Una vez elaborado el mapa de transcripción de los extremos del genoma donde se encuentran los mRNAs tempranos, se abordó la elaboración del mapa correspondiente al resto de los mRNAs existentes para así poder completar el mapa de transcripción de Cp-1. Estos mRNAs son los correspondientes a los genes implicados en la morfogénesis y en el proceso lítico, así como a los genes estructurales. El procedimiento seguido fue el mismo que se ha descrito en el apartado anterior, es decir, se construyeron sondas específicas de distintas regiones del genoma y se hibridaron con el RNA total, obtenido a diferentes tiempos, a partir de cultivos de neumococo infectados, o no, con Cp-1. Estos mRNAs se empezaron a detectar aproximadamente a los 35-40 min después de la infección, alcanzando

su máximo nivel de expresión a los 60 min; a partir de ese momento, la expresión va disminuyendo lentamente hasta que tiene lugar la lisis, como se ha señalado previamente. Las sondas utilizadas en estos experimentos aparecen en la Tabla 12, aunque sólo algunas de ellas están indicadas en la Figura 11A. Para no mostrar exhaustivamente todos los "Northern-blot" realizados, sólo se comentan a continuación los que se consideran más necesarios para la comprensión de la elaboración del mapa.

En la hibridación realizada con la sonda II, que contiene la ORF9, 10 y parte de la 11, se detectaron cinco bandas, aunque a los 30 min después de la infección (calle 1) aparecen esencialmente dos bandas de 2,6 y 1,8 kb (mRNAs H e I, respectivamente) (Figura 11A). A los 60 min (calle 2) se observaron además unas bandas de 6,8 y 5,2 kb mientras que, muy débilmente, se pudo detectar una banda de 11 kb. Si se tiene en cuenta la localización de los promotores (Tabla 11) y los posibles terminadores de la transcripción (Figura 5), la banda prominente de 1,8 kb corresponde al mRNA transcrito de las ORF8 y 9. Por otra parte, la sonda III, que comprende la ORF15 y parte de la 16, hibridó aparentemente sólo con dos bandas de 6,8 y 3,4 kb (Figura 11A). No obstante, si consideramos la localización de los promotores y terminadores, la banda de 6,8 podría representar un doblete compuesto por un mRNA de 6,8 kb que comienza en la ORF8, y un mRNA de 6,7 kb que comienza en la ORF12 (mRNAs F y J, respectivamente). El rastro que aparece por debajo del doblete podría corresponder a la banda de 5,3 kb (mRNA G) que solapa parcialmente (20 nt) con la sonda III. Si se comparan las bandas de hibridación detectadas con las sondas II y III, se puede postular la presencia de cuatro mRNAs, E, F, G e I que comienzan en el mismo promotor en tándem PL1/PL2. Estos datos se pueden relacionar con los niveles de proteínas sintetizadas en la célula infectada, particularmente de los productos de los genes situados en los extremos 5' de dichos mRNAs. Como resultado se puede concluir que, las proteínas codificadas por la ORF8, presunta proteína de andamiaje, y la ORF9, la proteína mayoritaria de la cabeza del fago, deben ser los productos más abundantes, aunque la proteína de andamiaje no esté presente en el virión, como se ha demostrado en otros sistemas (Casjens y Hendrix, 1988). Por otro lado, cuando se utilizó la sonda IV en los experimentos de "Northern-blot" aparecieron dos bandas de 3,3 y 1,6 kb (Figura 11A). La banda de mayor tamaño se corresponde con un mRNA que comienza en el promotor PL5 y acaba en el extremo 3' de la ORF19 (mRNA L). La mayor intensidad de la banda más pequeña, la de 1,6 kb, se podría deber a la presencia de dos mRNAs, uno que

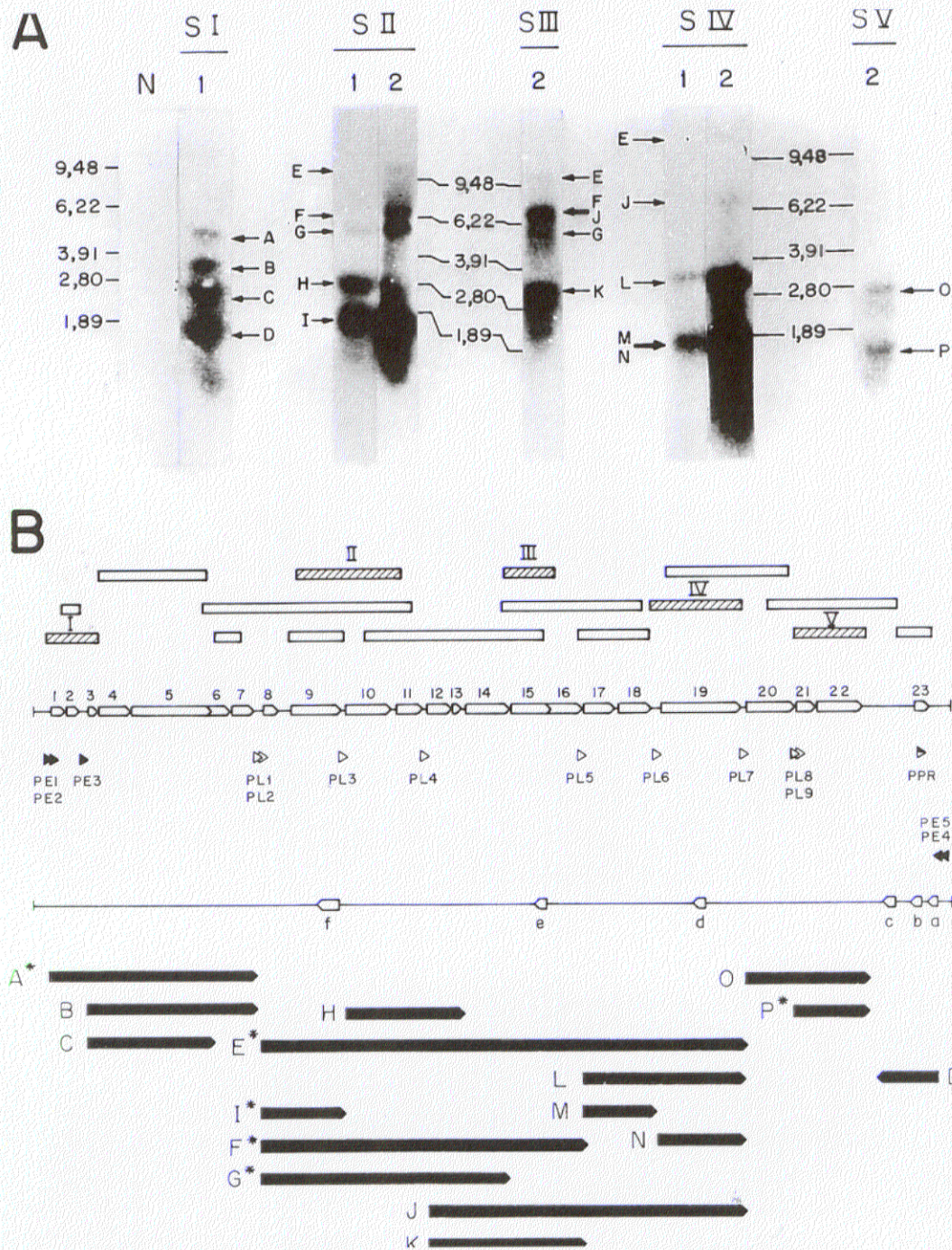


Figura 11. Transcripción del genoma del fago Cp-1. A) "Northern-blot". El RNA total se aisló de las células de neumococo sin infectar (calle N), a los 30 min (calles 1) o a los 60 min (calles 2) después de la infección. Las membranas se hibridaron utilizando como sonda el DNA de Cp-1 (calle N) o las sondas que aparecen en el panel B. Se cargaron aproximadamente 10 µg de RNA en cada calle. Las flechas indican los diferentes mRNAs detectados y se corresponden con los representados en el panel B. Las flechas gruesas representan mRNAs dobles. Los marcadores de peso molecular se representan en kb. B) Esquema del mapa de transcripción de Cp-1. Las ORFs se representan en flechas blancas en las dos cadenas del DNA indicando la dirección de la transcripción. Algunas de las sondas empleadas para los experimentos de "Northern-blot" se muestran como cajas blancas y las utilizadas en el panel A están indicadas en cajas rayadas. Los promotores tempranos y tardíos se indican con puntas de flechas negras y blancas, respectivamente; el promotor PPR se señala con una punta de flecha blanca y negra. Los mRNAs están marcados como flechas negras precedidas por letras mayúsculas. Los asteriscos indican mRNAs que utilizan promotores en tándem (ver texto para más detalles).

comenzaría en el promotor PL5 y acabaría en el presunto terminador localizado en el extremo 3' de la ORF18 (Figura 5) (mRNA M), y el mRNA N que se transcribiría desde el promotor PL6 terminando en el mismo sitio del mRNA L. La débil banda de 6,7 kb correspondería al mRNA que se transcribe desde el promotor PL4 y cuyo final de transcripción es el mismo que el de los mRNAs L y N. Finalmente, con la sonda V sólo se detectaron dos bandas de 2,6 y 1,4 kb, en cada uno de los tiempos analizados (Figura 11A). Estos mRNAs corresponden a los RNAs transcritos que comienzan en el promotor PL7 (mRNA O) y desde los promotores en tándem PL8/PL9 (mRNA P) y que usan como señal de terminación de transcripción la situada en el extremo 3' de la ORF22 (Figura 5). La baja intensidad de estas bandas se relaciona con el reducido nivel de expresión de los genes de encapsidación y de los implicados en la lisis.

Todos estos resultados referidos tanto a los genes tempranos como a los tardíos, permiten elaborar de forma esquemática un mapa de transcripción de Cp-1 que aparece en la Figura 11B.

2.4. RNA implicado en la encapsidación del DNA

En el análisis de promotores funcionales a partir de la secuencia de nt se localizó un posible promotor en la región 5' de la ORF23, próximo a la ITR de la derecha del genoma. Sin embargo, el experimento de "primer extension" reveló la existencia de una molécula de RNA que comenzaba 11 nt corriente abajo del codón inicial, ATG, de la ORF23 (Figura 12A), lo que descartaba que este RNA transcrito pudiera corresponder a la síntesis de alguna proteína. A continuación, se analizó la estructura secundaria de este RNA, puesto que se ha descrito la implicación de ciertas moléculas de RNA en diferentes funciones biológicas (King y cols., 1986). Como resultado de este análisis se obtuvo una predicción óptima de estructura valiéndose de una molécula de RNA de 169 nt, como puede verse en la Figura 12B. Esta estructura es muy similar a la encontrada para los RNAs de las precabezas de ϕ 29 y los fagos con él relacionados (PZA, GA1, M2, entre otros) (Figura 12C) (Bailey y cols., 1990), donde se ha demostrado su implicación en la encapsidación del DNA, a través de una reacción dependiente de ATP (Guo y cols., 1987d). En las secuencias de estos pequeños RNAs, también llamados pRNAs o RNAs pilotos, no se han encontrado similitudes significativas, aunque sí presentan una misma estructura secundaria (Bailey y

cols., 1990). En el caso del RNA de Cp-1 se pueden identificar los dos dominios I y II, presentes en los pRNAs ya descritos. En el dominio I se pueden distinguir las cuatro hélices A, B, C, D, y en el dominio II, la hélice E y la F (Figura 12B). Las diferencias entre los pRNAs estriban en la longitud de las hélices y la extensión de las zonas no apareadas, aunque tales variaciones son mínimas, pues la estructura secundaria global es la misma para todos ellos. La implicación de un factor no proteico en el empaquetamiento del DNA viral aporta nuevos datos a las interacciones RNA/DNA y RNA/proteínas y aumenta las funciones previamente demostradas del RNA (Inouye y Delihás, 1988).

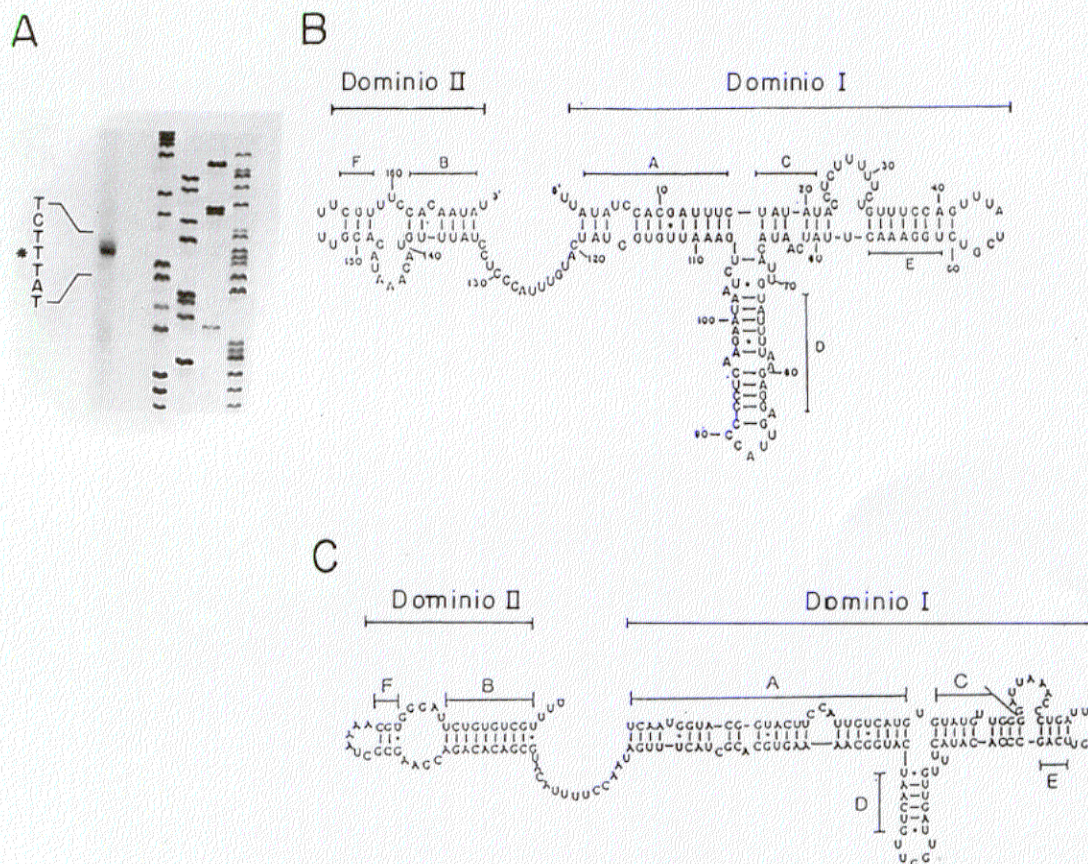


Figura 12. Identificación del pRNA de Cp-1. A) "Primer extension". Mapeo del extremo 5' del correspondiente RNA. El método usado fue similar al descrito en la Figura 9. B) Estructura secundaria del pRNA de Cp-1. C) Estructura secundaria del pRNA de $\phi 29$. En B y C se indica la localización de los dos dominios y las hélices de acuerdo con Bailey y cols. (1990).

3. ANÁLISIS DE LAS ORFs DEL DNA DE Cp-1

3.1. Clonación y expresión de la ORF4 que codifica la TP y de la ORF5 que codifica la DNA polimerasa

La DNA polimerasa y la TP son, como ya se ha comentado, proteínas esenciales en el proceso de replicación del DNA, que se lleva a cabo por el mecanismo peculiar de iniciación con proteína o “protein-priming”. Por ello, nos propusimos clonar y expresar las ORFs 4 y 5 que codifican la TP y la DNA polimerasa, respectivamente, como paso previo para profundizar en el conocimiento del mecanismo de replicación del genoma del fago.

En primer lugar se consiguió clonar dichas ORFs en los vectores de expresión pIN-III-A3 y pT7-4, utilizando para ello sitios de restricción apropiados. No obstante, no se consiguió detectar las proteínas correspondientes en geles de poliacrilamida-SDS teñidos con azul de Coomassie por lo que se intentó una nueva estrategia. Para ello, se amplificaron los fragmentos correspondientes de los genes de la DNA polimerasa y la TP por medio de la técnica de PCR. Los oligonucleótidos sintéticos utilizaban un sitio de restricción *NdeI* en el extremo 5' del DNA amplificado y otro *EcoRI* en el extremo 3'. El fragmento de DNA amplificado se trató con fenol, se precipitó con etanol, se digirió con *NdeI* y *EcoRI*, se purificó por la técnica de la β -agarasa, se ligó al vector pT7-7 previamente digerido con las mismas enzimas y se transformó la cepa K514 de *E. coli*. De esta manera, se obtuvieron dos plásmidos, pAMR50, que contenía el gen de la DNA polimerasa, y pAMR40, que contenía el de la TP (Figura 13).

Para determinar si los plásmidos recombinantes codificaban las proteínas esperadas se analizó su expresión en maxicélulas de *E. coli*. Para ello, la cepa CM12 se transformó con pAMR40, pAMR50 o pT7-7 como control. Se obtuvieron los extractos celulares siguiendo el protocolo descrito en el apartado 17.1 de Materiales y Métodos y posteriormente se analizaron en un gel de poliacrilamida-SDS al 10%. Como se puede comprobar en la Figura 14 los plásmidos recombinantes sintetizaron proteínas con la movilidad electroforética esperada, pues en el caso de pAMR40 se detectó una proteína de 29 kDa, y en el caso de pAMR50 la proteína sintetizada es de 64 kDa, que se correlacionaban bien con las masas moleculares esperadas para la TP y la DNA polimerasa, respectivamente. Por otra parte, se analizó la producción de estas proteínas en geles de

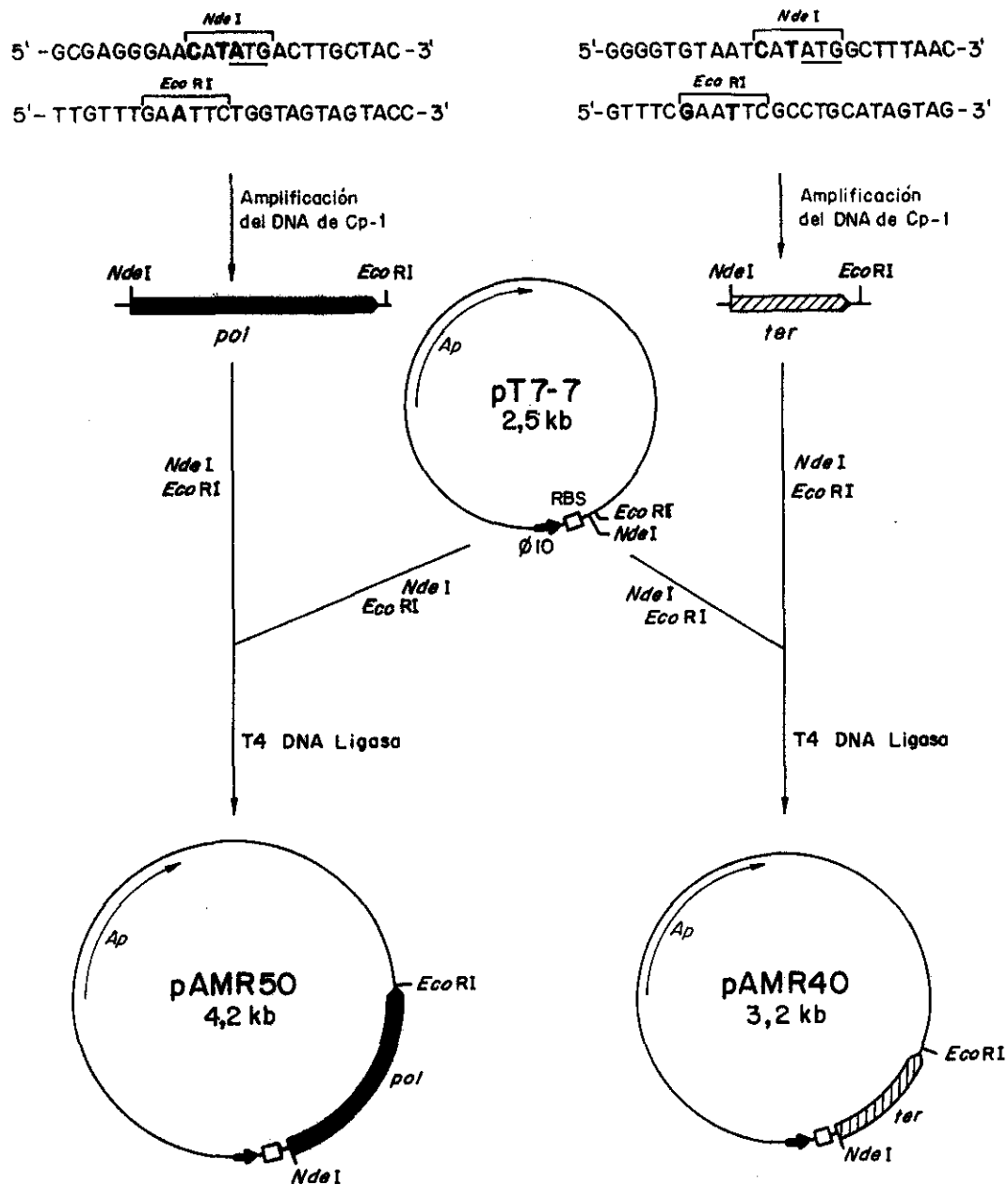


Figura 13. Construcción de pAMR40 y pAMR50. En la parte superior se indican los oligonucleótidos utilizados para amplificar el DNA de Cp-1 por la técnica de PCR. En negrita aparecen los nucleótidos mutados con respecto a la secuencia del DNA de Cp-1, para crear los sitios de restricción indicados en cada caso. Los codones de iniciación aparecen subrayados. La flecha negra gruesa representa la ORF5 que codifica la DNA polimerasa (pol) y la flecha rayada la ORF4 que codifica la TP (ter). El cuadrado blanco representa el RBS aportado por el plásmido. La flecha negra pequeña indica el promotor ϕ 10 del fago T7.

poliacrilamida-SDS teñidos con azul de Coomassie. Para ello, se transformó la cepa BL21(DE3) de *E. coli* con los plásmidos anteriores, obteniéndose los extractos crudos como se describe en el apartado 16.1 de Materiales y Métodos. La Figura 15 pone de manifiesto que el extracto correspondiente al pAMR40 contenía una proteína adicional, en comparación con la muestra control del pT7-7, que tiene el mismo peso molecular deducido para la TP. De igual manera, en el extracto procedente del pAMR50 se detectó una proteína suplementaria de 64 kDa que coincide con la masa molecular de la DNA polimerasa.

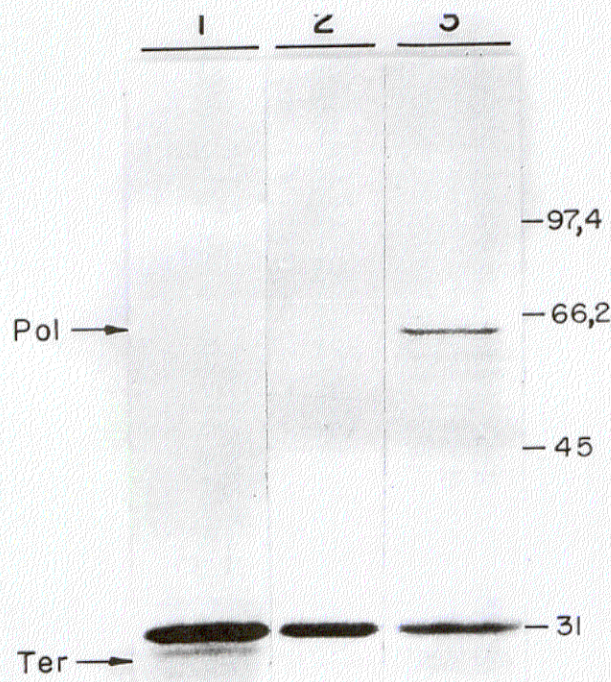


Figura 14. Máxicélulas de CM12(pT7-7), CM12(pAMR50) y CM12(pAMR40). Extractos de cultivos de CM12(pAMR40) (calle 1), CM12(pT7-7) (calle 2) y CM12(pAMR50) (calle 3) se cargaron en un gel de poliacrilamida-SDS del 12%. Las flechas indican la posición de la DNA polimerasa (Pol) y la proteína terminal (Ter). Los marcadores de peso molecular, en kDa, se muestran a la derecha.

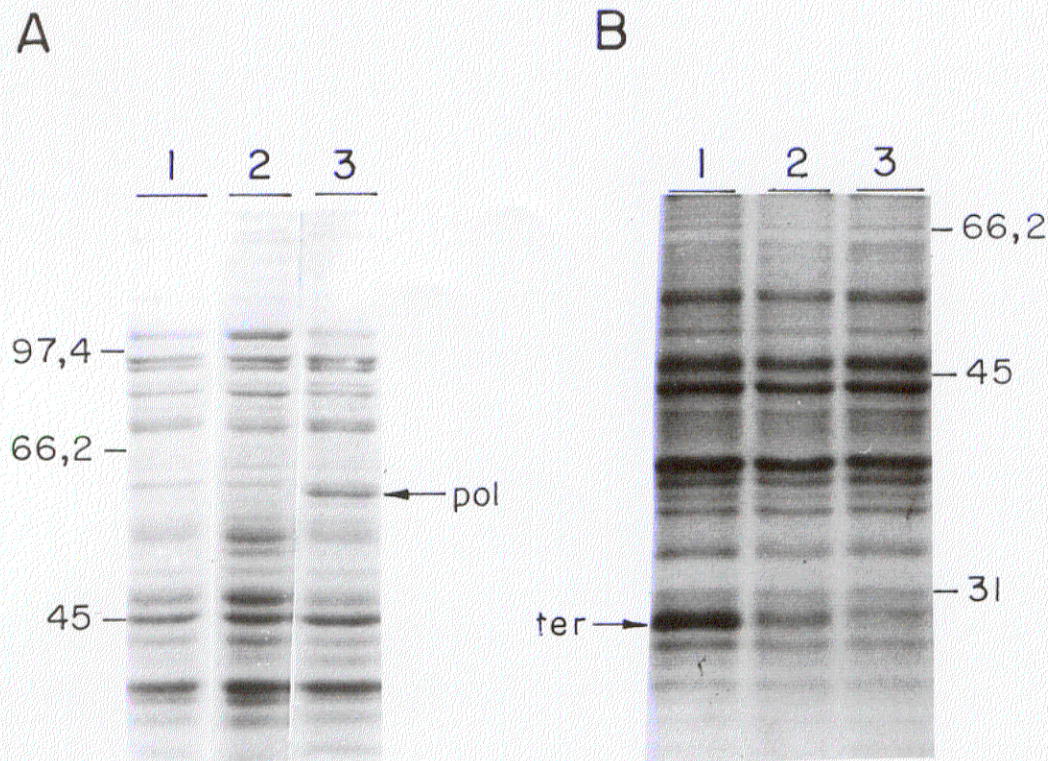


Figura 15. Expresiones de las ORFs 4 y 5 en *E. coli*. A) Expresión del gen que codifica la DNA polimerasa de Cp-1. Extractos de cultivos de BL21(DE3)(pT7-7) inducidos con IPTG (calle 1), BL21(DE3)(pAMR50) sin inducir (calle 2) y BL21(DE3)(pAMR50) inducidos con IPTG (calle 3) se cargaron en un gel de poliacrilamida-SDS del 10%. La flecha indica la posición de la DNA polimerasa (pol). Los marcadores de peso molecular, en kDa, se muestran a la izquierda. B) Expresión del gen que codifica la TP de Cp-1. Extractos de cultivos de BL21(DE3)(pAMR40) inducidos con IPTG (calle 1), BL21(DE3)(pAMR40) sin inducir (calle 2) y BL21(DE3)(pT7-7) inducidos con IPTG (calle 3) se cargaron en un gel de poliacrilamida-SDS del 10%. La flecha indica la posición de la TP (ter). Los marcadores de peso molecular, en kDa, se muestran a la derecha.

En lo que se refiere a ensayos funcionales de estas proteínas esenciales para la replicación del DNA, no se ha podido demostrar la actividad DNA polimerasa de los extractos de BL21(DE3)(pAMR50) utilizando ensayos de polimerización *in situ* en geles de poliacrilamida-SDS (Karawya y cols., 1983), ni tampoco ensayos de iniciación de la replicación *in vitro* con extractos de BL21(DE3) transformada con dichos plásmidos.

3.2. Determinación de la secuencia N-terminal de las proteínas codificadas por las ORFs 1, 2, 6, 7, 10, 15, 16 y 17 en *E. coli*

Con objeto de confirmar que determinadas ORFs del genoma de Cp-1 codificaban las proteínas esperadas, nos propusimos analizar los residuos N-terminales de las proteínas fágicas producidas en *E. coli*. Para ello, se utilizaron los plásmidos recombinantes descritos en el apartado 1.1 de Resultados y en los cuales el inserto del DNA fágico estuviera en la dirección favorable con respecto al promotor del gen *lacZ*. Con los plásmidos pGG2, pGG57, pGG61, pGG66 y pAMR45 se transformó la cepa DH5 α y se establecieron las condiciones de máxima producción de las proteínas buscadas según se describe en el apartado 16.1 de Materiales y Métodos.

Los geles de poliacrilamida-SDS mostraron unas claras bandas de proteínas en las posiciones esperadas para cada extracto de la cepa transformada. De esta manera, se consiguió asignar cada banda mayoritaria a la proteína buscada y, siguiendo la metodología descrita en el apartado 21 de Materiales y Métodos, se analizaron los residuos N-terminales de dichas proteínas (Tabla 13).

Tabla 13. Secuencia N-terminal de proteínas de Cp-1 sintetizadas en *E. coli*^a

Plásmido	ORF ^b	Secuencia
pGG61	1	MKMKAL
pGG61/pGG66	2	MELRQL
pAMR45	6	MKQPLKSE
pAMR45	7	AKQAPKH
pGG57	10	SYKNY
pGG2	15	MNPEEFKD
pGG2	16	PLETRS
pGG2	17	MQUESTKI

^a Los cultivos de DH5 α transformados con los correspondientes plásmidos se multiplicaron como se detalla en el apartado 16.1 de Materiales y Métodos.

^b Indica la ORF cuyo producto se analizó según se explica en el apartado 21 de Materiales y Métodos.

Como se puede observar, en los productos codificados por las ORFs 7, 10 y 16 se procesa la metionina inicial. El procesamiento en estas proteínas y no en las otras, corrobora la teoría de que la probabilidad de procesamiento de la metionina inicial disminuye a medida que aumenta el tamaño del aminoácido situado en segunda posición (Hirel y cols., 1989).

Asimismo, con estos mismos extractos se hicieron experimentos de "Western-blot" con antisuero anti-Cp-1, detectándose una banda de reconocimiento para las proteínas correspondientes a las ORFs 10, 15, 16 y 17. Por análisis comparativo de secuencia se había deducido que las ORFs 10 y 17 codifican las proteínas del conector y la cola, respectivamente. Sin embargo, no se conoce la función de las ORFs 16 y 17, pero el resultado positivo de "Western-blot" junto con su localización en el genoma y su expresión en tiempos tardíos se puede postular que codifican proteínas estructurales. Por el contrario, las proteínas codificadas por las ORFs 1, 2, 6 y 7 no son reconocidas por el antisuero, como cabría esperar, ya que al ser proteínas tempranas no formarían parte de la estructura del virión sino que, como ya se ha señalado, estarían implicadas en procesos de replicación o regulación de la transcripción.

3.3. ESTUDIOS SOBRE LA PROTEÍNA MAYORITARIA DEL VIRIÓN, PRODUCTO DE LA ORF9

3.3.1. Procesamiento de la proteína mayoritaria del virión

En los primeros geles de poliacrilamida-SDS en los que se analizaron las proteínas del virión se observó la presencia de una proteína mayoritaria que constituía aproximadamente el 90% de las proteínas totales y presentaba un masa molecular de 37 kDa (López y cols., 1984). Para asignar esta proteína a una ORF del genoma del fago se aisló la banda del gel y se determinó la secuencia de su N-terminal. La secuencia obtenida, ATVPYQDFQ, no se correspondía con los primeros aminoácidos de ninguno de los productos de las ORFs, sino con la secuencia que empieza en el aminoácido 49 del producto de la ORF9. Como ya se ha descrito en el apartado 1.3 de Resultados, la ORF9 codifica probablemente la proteína mayoritaria de la cabeza, dada la homología de secuencia que presenta con la proteína mayoritaria de la cabeza de ϕ 29. Por tanto, la secuencia N-terminal obtenida indicaba que esta proteína sufría una proteólisis específica. En este

sentido, el procesamiento de las proteínas que forman la cabeza de los viriones, así como el de las diferentes proteínas que intervienen en el ensamblaje del fago, no es un hecho infrecuente sino que se han descrito y caracterizado con detalle muchos ejemplos en distintos fagos (Casjens y Hendrix, 1988; Duda y cols., 1995a, b; Myung y cols., 1994; Loessner y Scherer, 1995)

Para obtener datos precisos sobre el procesamiento de esta proteína así como sobre el origen de la proteasa, nos propusimos expresar la proteína mayoritaria en *E. coli* y determinar si también se procesaba en este sistema. Para construir el plásmido recombinante pAMR61, que contiene la ORF9, se partió del plásmido pGG9 (Tabla 5), y se procedió como se describe en la Figura 16. Con este plásmido se transformó DH5 α y de las colonias correspondientes se obtuvieron extractos, comprobándose en geles de poliacrilamida-SDS la aparición de una proteína de una masa molecular correspondiente a la deducida de la secuencia del gen que presuntamente codifica la proteína de la cabeza (42 kDa) (Figura 17).

Es interesante resaltar que en los cultivos de la cepa DH5 α (pAMR61) el nivel de expresión de la proteína fágica era muy alto incluso en ausencia de inductor y cuando el cultivo se observó al microscopio óptico se veían cuerpos de inclusión en los polos de las células. A partir de un gel preparativo de la muestra anterior, se secuenció el N-terminal de la proteína fágica corroborándose la hipótesis de que la proteína no se procesaba en *E. coli*, puesto que coincidía con los primeros aminoácidos deducidos a partir de la secuencia de la ORF9 excepto la metionina inicial, es decir ANKITTFLS. Esto sugería que el procesamiento de los 48 primeros aminoácidos en el N-terminal de la proteína es un proceso específico del sistema de neumococo, aunque no se podía especificar si la proteólisis la lleva a cabo una proteína codificada por el huésped o por el propio genoma fágico.

Con el fin de poder contestar a este interrogante, se preparó un antisuero contra la proteína mayoritaria del virión eluida del gel de poliacrilamida-SDS al 10%. De esta manera, para determinar cómo se producía el procesamiento *in vivo* durante la infección, los extractos obtenidos de cultivos de neumococo infectados con Cp-1 a diferentes tiempos después de la infección fueron analizados mediante "Western-blot". En la Figura 18A se puede observar que el procesamiento se produce tan pronto como se empieza a sintetizar la proteína, pero la eficiencia del proceso no es del 100%, ya que desde el primer momento aparecen las dos formas moleculares de la proteína. Además, la proporción entre ambas

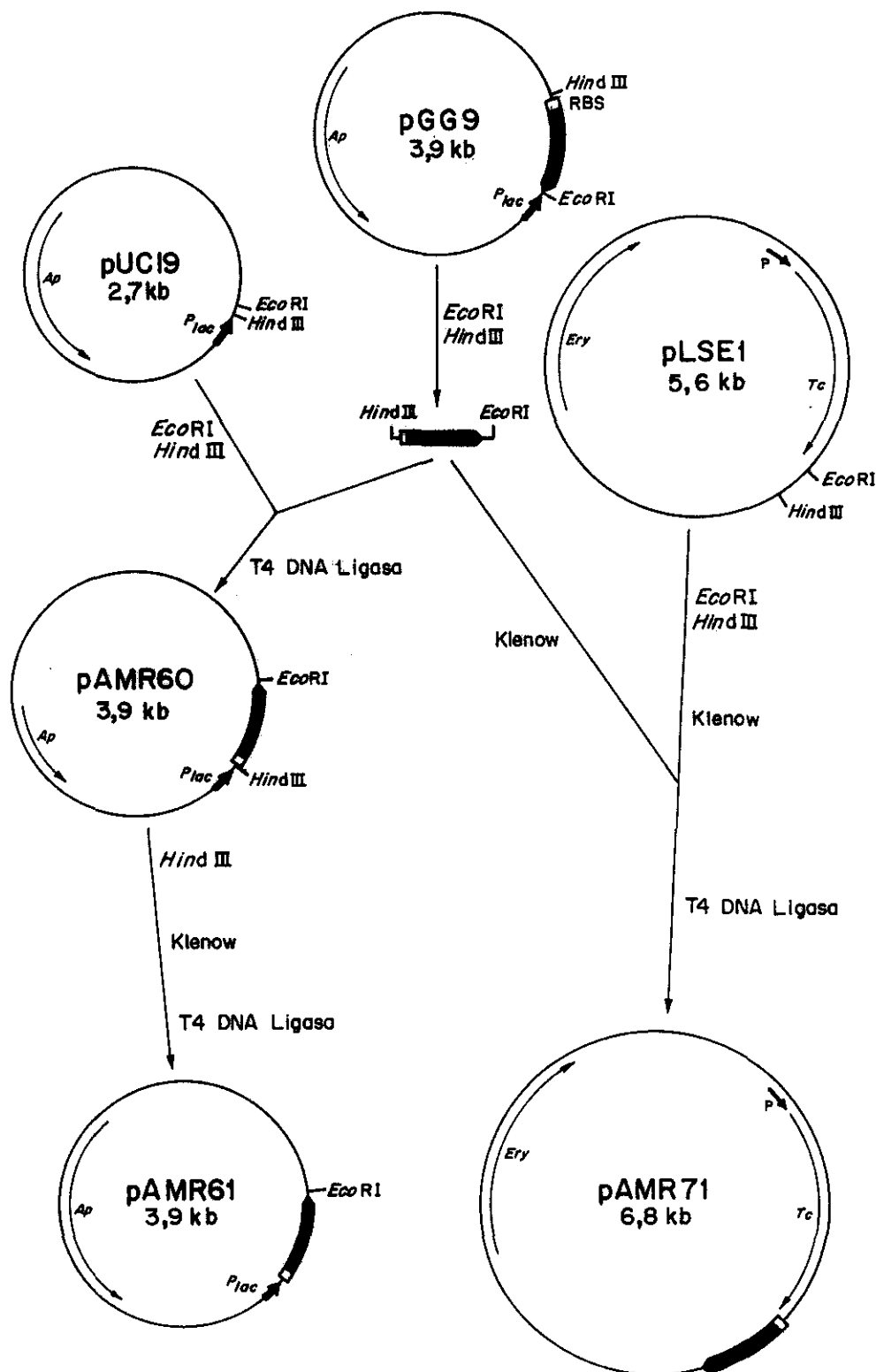


Figura 16. Construcción de pAMR61 y pAMR71. La flecha negra gruesa representa la ORF9 que codifica la proteína de la cabeza de Cp-1. El rectángulo blanco que precede a la ORF9 indica su RBS. P indica el promotor del gen de la resistencia a Tc y P_{lac} el del gen *lacZ*.

moléculas parece mantenerse constante durante la infección. Este dato contrasta con el hecho de que en el virión solo se encontró la proteína procesada (Figura 18A), lo que sugiere que la proteína sin procesar no es capaz de formar viriones infectivos. Por otra parte, ambas proteínas se encuentran en la banda que posee menor densidad de flotación que la correspondiente al virión y que se obtiene en la primera centrifugación en el gradiente de CsCl para la purificación de Cp-1 (Figura 18A). Esta banda, que no contiene DNA, estaría formada por restos de proteínas fágicas, y de viriones defectuosos incapaces de empaquetar el DNA y formar un virión infectivo.

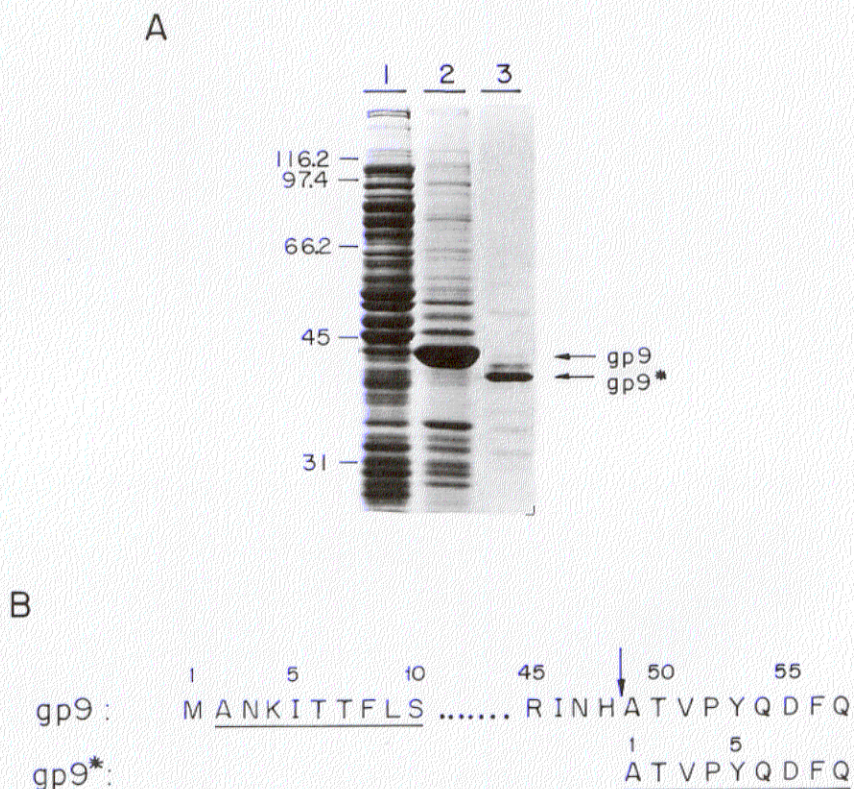


Figura 17. Expresión de la proteína mayoritaria de la cabeza del fago Cp-1. A) Extractos de cultivos de DH5 α (pUC19) (calle 1), DH5 α (pAMR61) (calle 2) o del virión Cp-1 purificado en gradiente de CsCl (calle 3) se cargaron en un gel de poliacrilamida-SDS del 10%. Los marcadores de peso molecular se indican en kDa. B) Secuencias de aminoácidos del producto de la ORF9 (gp9) y de su forma procesada (gp9*). La flecha vertical indica el sitio de corte y los aminoácidos subrayados son los obtenidos del análisis de su N-terminal. Los números de los aminoácidos de gp9 se indican de acuerdo con la secuencia de nucleótidos de la ORF9.

Como paso previo para determinar si la proteasa está codificada por el fago o por el huésped se construyó el plásmido pAMR71 que deriva del pLSE1 y que contiene la ORF9 que codifica la proteína de la cabeza, un plásmido que se puede replicar en *E. coli* y en neumococo, y que no posee promotores inducibles. La construcción del pAMR71 se realizó como se describe en la Figura 16, y hay que resaltar que la expresión de la ORF9 queda bajo el control del promotor constitutivo del gen que codifica la proteína que confiere resistencia a la tetraciclina. Previamente a la transformación de la cepa M31 de neumococo se comprobó la construcción en la cepa C600 de *E. coli*, y a continuación se transformó M31 con pAMR71 o con pLSE1 como control. Los extractos crudos de los transformantes correspondientes se analizaron en geles de poliacrilamida-SDS al 10% y, teñidos con azul de Coomassie, no se aprecia diferencia entre ambas muestras, lo que era de esperar dado el sitio de clonación de este gen. Por tanto, para poder determinar el peso molecular de la proteína expresada en neumococo se hicieron experimentos de “Western-blot” utilizando los anticuerpos obtenidos contra la proteína mayoritaria del virión. Dado el bajo nivel de expresión de la ORF9 en *S. pneumoniae* se partió de 100 ml de cultivo de estas cepas transformadas para la obtención de los extractos correspondientes. Como se puede ver en la Figura 18B, se detectó por “Western-blot” una banda a la altura de la proteína sin procesar en el extracto de células que contiene el plásmido pAMR71 y que no apareció en el que contiene el pLSE1. Por el contrario, no aparecía ninguna banda a la altura de la proteína procesada, presente en el virión, lo que indicaba, o bien que la proteasa estaba codificada por el fago o que la proteasa estaba codificada por el huésped pero sólo se expresaba en determinadas condiciones de stress como pueden ser la infección por fagos, tratamiento con radiación UV, mitomicina, etc. Para tratar de resolver este dilema, se prepararon extractos de M31(pAMR71) y M31(pLSE1) en diferentes condiciones: infectados con Dp-1, tratados con mitomicina o con UV. Como se muestra en la Figura 18B, en todos los extractos sólo aparecía una banda a la altura de la proteína sin procesar que no estaba presente en los extractos controles. Del resultado de esta experimento se pudo concluir o que la proteasa está codificada por el fago, o que éste sintetiza algún factor que induce al huésped a producirla.

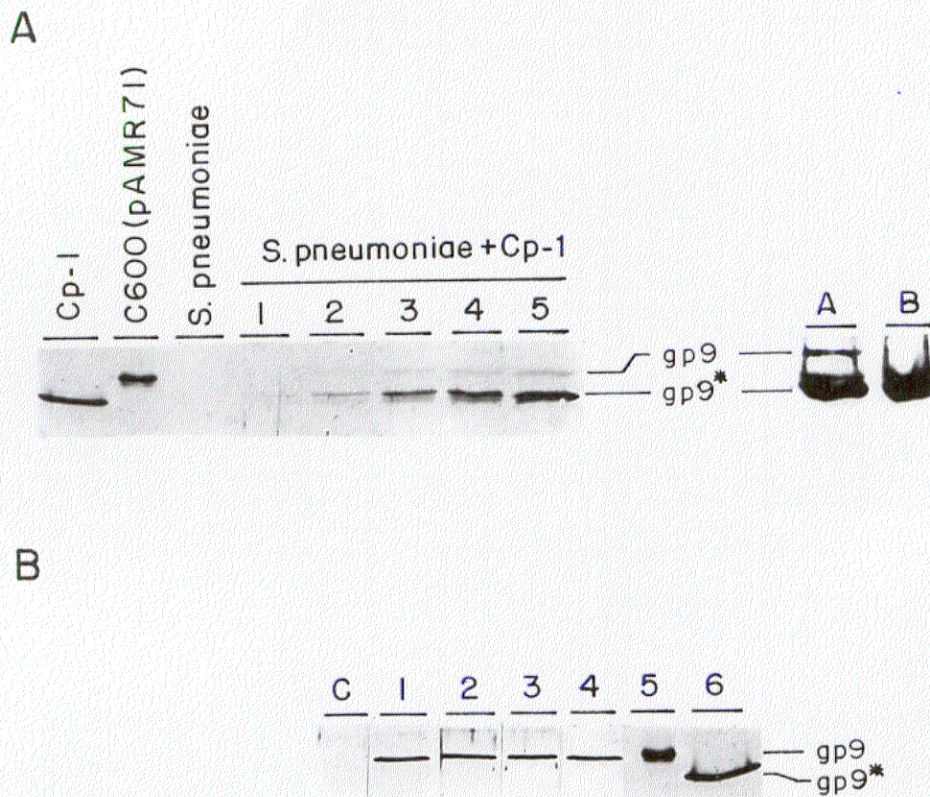


Figura 18. Detección de la proteína mayoritaria de la cabeza por experimentos de “Western-blot”. **A)** En la parte izquierda se analizó el virión Cp-1, extractos de C600(pAMR71), extractos de neumococo sin infectar e infectados, tomados a 40 (calle 1), 60 (calle 2), 80 (calle 3), 100 (calle 4) y 120 min (calle 5) después de la infección. En la parte derecha se analizó la banda superior (calle A) o la del virión (calle B) del bandeo en CsCl. **B)** Extractos de M31(pLSE1) (calle C), M31(pAMR71) (calle 1), M31(pAMR71) irradiado con UV (calle 2), M31(pAMR71) con mitomicina C (calle 3), M31(pAMR71) infectado con Dp-1 (calle 4), C600(pAMR71) (calle 5) y el virión Cp-1 (calle 6). Todas las muestras se cargaron en un gel de poliacrilamida-SDS del 10% y se analizaron con el anticuerpo anti-gp9*. Se indican las posiciones de gp9 y de su forma procesada, gp9*.

3.3.2. La proteasa que procesa la proteína mayoritaria del virión está codificada por el propio fago

Para comprobar la hipótesis del posible origen fágico de la proteasa se estudió el procesamiento de la proteína en cultivos de *E. coli* que expresan el gen de la proteína de la cabeza con la expresión de otras ORFs codificadas por el fago. De las 29 ORFs que contiene el genoma de Cp-1, 11 de ellas codifican proteínas cuyas secuencias presentaban un alto grado de similitud con proteínas de función conocida del fago ϕ 29, lo que permitía atribuirles una posible función (apartado 1.3 de Resultados), y ninguna parece corresponder a una posible proteasa. De las 18 restantes se podían descartar las que se expresan en periodos tempranos puesto que normalmente codifican proteínas implicadas en replicación o regulación de la transcripción. Por consiguiente, nos centramos en las proteínas cuyos genes se expresan tardíamente y, además, no tienen posible función asignada, es decir, las codificadas por las ORFs 12 a 16 y 18.

Por otra parte, para determinar la secuencia del genoma se habían construido plásmidos que contenían diferentes ORFs del genoma (Tabla 5). Así, pGG1, contenía las ORFs 5 a 14 completas además de la parte C-terminal de la ORF4 y la parte N-terminal de la ORF15 (Figura 19A). Los “Western-blot” realizados con el extracto de la cepa DH5 α (pGG1) de *E. coli*, a diferencia de los obtenidos a partir de *E. coli*(pAMR61) que solo contenía la ORF9, mostraban una banda de 37 kDa a la altura de la correspondiente a la proteína procesada del virión (Figura 19B). Este resultado indicaba que alguna de las proteínas codificadas por el plásmido pGG1 era la responsable del procesamiento de la proteína.

En otro experimento complementario de “Western-blot” se probó el extracto preparado de DH5 α (pAMR45). Este plásmido contiene las ORFs 6 a 10 completas, es decir dos ORFs cuyos genes son tempranos (ORF6 y 7), la ORF9 que corresponde a la proteína mayoritaria y otras dos ORFs que la flanquean (ORF8 y 10). El resultado puso de manifiesto la detección de una banda a la altura de la proteína sin procesar, por lo que, en principio, sería descartable que la proteasa estuviera codificada por una de las ORFs presentes en el pAMR45 (datos no mostrados).

Con los resultados anteriores, las ORFs que podrían codificar la proteasa quedaban reducidas a las ORFs 12, 13 y 14, puesto que la ORF15 no estaba completa. De todas

formas, podría suceder que el producto de la ORF15 fuera activo sin la región C-terminal, por lo que para eliminar esta posibilidad se transformó C600(pAMR71) con el plásmido pGG2, que como vimos en el apartado 3.2 de Resultados sí expresa la ORF15. El resultado obtenido por “Western-blot” con este extracto fue negativo, es decir, no se trataba de la proteasa porque sólo se detectaba la proteína sin procesar (Figura 19B).

Por otra parte, se transformó la cepa C600(pAMR71) con el plásmido pGG6 que contiene las ORFs 11 a 14 completas (Figura 19A) y en la misma orientación que el promotor del gen *lacZ*. El extracto correspondiente se analizó también por “Western-blot” detectándose en esta ocasión dos bandas, una a la altura de la proteína procesada y otra en la posición de la proteína sin procesar (Figura 19B). Este resultado parece indicar que la proteína mayoritaria de Cp-1 puede procesarse por alguna proteasa codificada por pGG6. El inserto fágico de pGG6 está contenido en el de pGG1, sin embargo, como ya se ha citado, el “Western-blot” del extracto de pGG1 muestra una sólo banda que corresponde a la proteína procesada y el de pGG6, dos bandas que corresponden a la proteína procesada y sin procesar. Esta aparente contradicción podría deberse a que en pGG1, la ORF9 que codifica la proteína de la cabeza no está hiperexpresada y está en *cis* con respecto a la posible ORF que codifica la proteasa. Por el contrario, en el caso del extracto de C600(pAMR71)(pGG6), la ORF9 está contenida en el plásmido pAMR71 (por tanto está en *trans* con respecto a la ORF contenida en pGG6 que codifica la posible proteasa) y el nivel de expresión es mucho mayor que en el pGG1. Esto parece indicar que la proteasa sólo es capaz de procesar en *E. coli* cierta cantidad de proteína de la cabeza cuando ésta se está sintetizando en gran cantidad. Como la ORF11 codifica presuntamente la proteína del collar del fago, cualquiera de las otras tres podía codificar la proteasa.

Para intentar asignar la función proteolítica a una de estas ORFs, aprovechamos la existencia de un corte de restricción adecuado para inactivar la ORF14. El plásmido pGG6 contiene un sitio *StyI* en el extremo 5' de la ORF14 y otro en el extremo 3' de la ORF15 (Figura 19A). Por tanto, se digirió el plásmido pGG6 con *StyI*, se purificó el fragmento mayor de la digestión y se religó, obteniéndose el plásmido pGG1213 (Figura 19A). Con este nuevo plásmido recombinante se transformó C600(pAMR71) analizándose un extracto del mismo por “Western-blot”. Como se muestra en la calle de la Figura 19B, sólo se detecta la banda que corresponde a la proteína sin procesar, por lo que se podría concluir que la ORF14 codifica la proteasa responsable del procesamiento de la proteína de la

cabeza. Para obtener una demostración directa de esta actividad proteolítica, recientemente se ha clonado la ORF14 en diferentes plásmidos de expresión, con los que se ha transformado la cepa de *E. coli* que expresa el gen de la proteína mayoritaria del virión. Aunque los resultados son todavía preliminares, parecen corroborar el papel de la ORF14 como proteasa fágica.

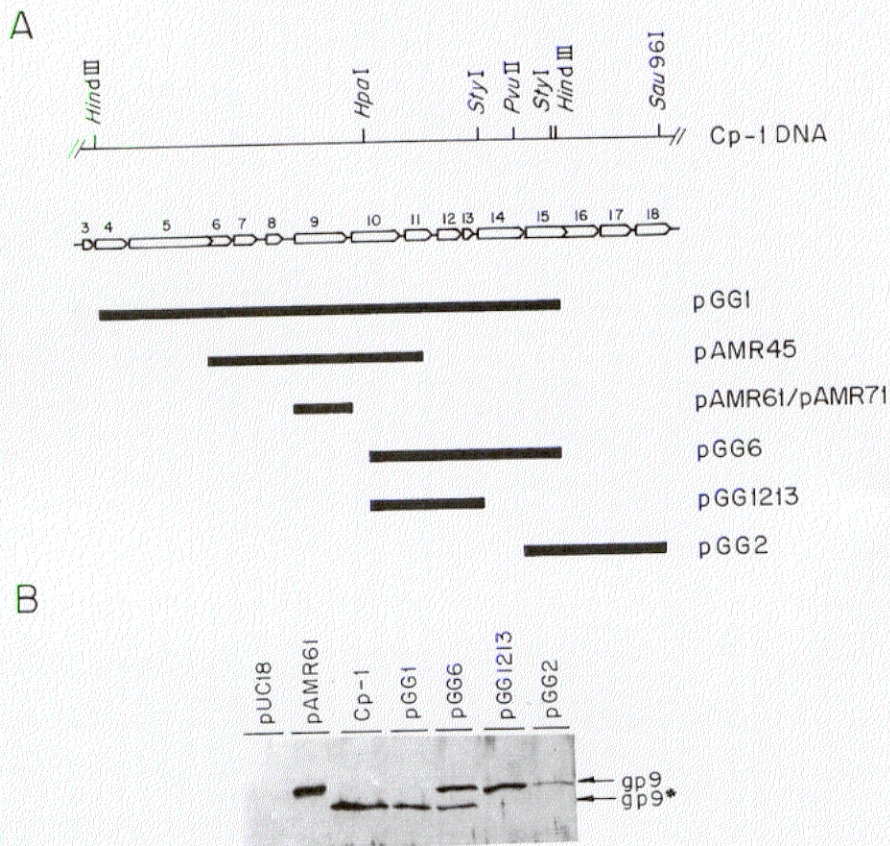


Figura 19. "Western-blot" de extractos de *E. coli* con anticuerpos anti-gp9*. **A)** Localización de los insertos fágicos de los plásmidos recombinantes, representados en cajas negras, que se han usado en los experimentos de "Western-blot". Las ORFs del genoma se indican en flechas blancas. **B)** Extractos de cepas de *E. coli* transformados con los plásmidos señalados, así como el virión de Cp-1, se cargaron en un gel de poliácridamida-SDS del 10% y se analizaron con el anticuerpo anti-gp9*. Se indican las posiciones de gp9 y de su forma procesada, gp9*.

3.4. ANÁLISIS DE LAS ORFs IMPLICADAS EN EL PROCESO DE LISIS

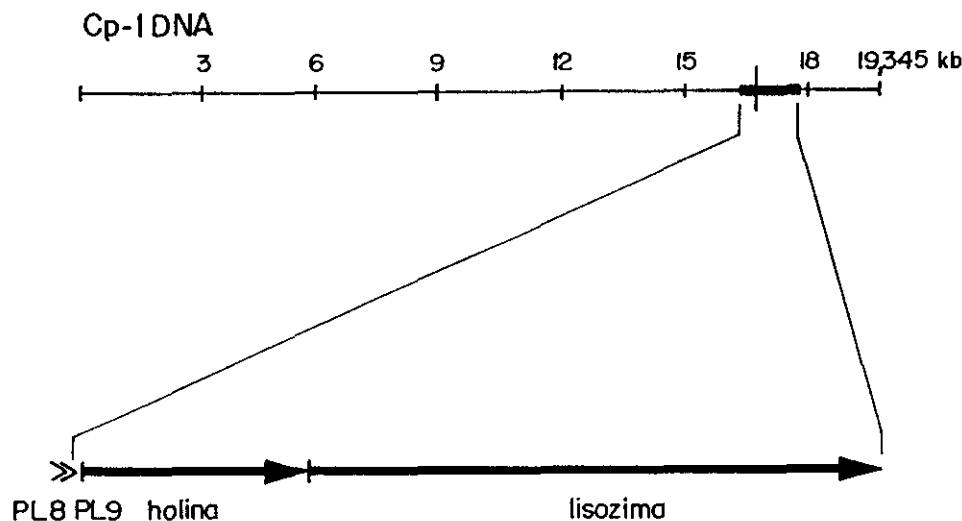
La enzima lítica de Cp-1 codificada por el gen *cpl1* (ORF22) se ha caracterizado como una lisozima dependiente de colina para su actividad (García y cols., 1987), y presenta una organización modular en la que el dominio N-terminal es responsable de la actividad lítica y el C-terminal de la unión a colina (Figura 20C) (García y cols., 1990). Precediendo a este gen se encuentra la ORF21, que codifica una proteína con todas las características propias de las holinas, como se ha comentado en el apartado 4.1.1 de la Introducción. Esto sugería un sistema dual de lisis en el fago Cp-1. Como se puede ver en la Figura 20B, esta holina, Cph1, no presenta el doble motivo de iniciación de la traducción que muestran otras holinas, el extremo C-terminal es muy hidrofílico con muchos aminoácidos cargados, aunque la carga neta es nula, y el extremo N-terminal presenta una carga neta positiva. Además, mediante la predicción de estructura secundaria se pueden postular tres regiones transmembrana separadas por giros β . El primer giro β , situado entre la primera y la segunda región transmembrana, se predice con alta fiabilidad, mientras que el segundo, situado entre la segunda y la tercera región transmembrana, no se puede predecir con tanta fiabilidad. Su punto isoeléctrico es 7,87. Los genes líticos se sitúan entre los nucleótidos 16.027 y 17.448 del genoma de Cp-1, solapando el codón de terminación de la holina con el codón de iniciación de la lisozima (característica presente en muchos de los genes líticos de los fagos) (Figura 20A). En el extremo 5' de estos genes se encuentra una región promotora con dos promotores en tándem, separados por 78 nt, como se ha demostrado en el apartado 2.1 de Resultados.

Para profundizar en el mecanismo de lisis del fago y demostrar la funcionalidad de la holina propuesta nos propusimos clonar y expresar estos genes líticos, tanto en el sistema heterólogo de *E. coli* como en el homólogo.

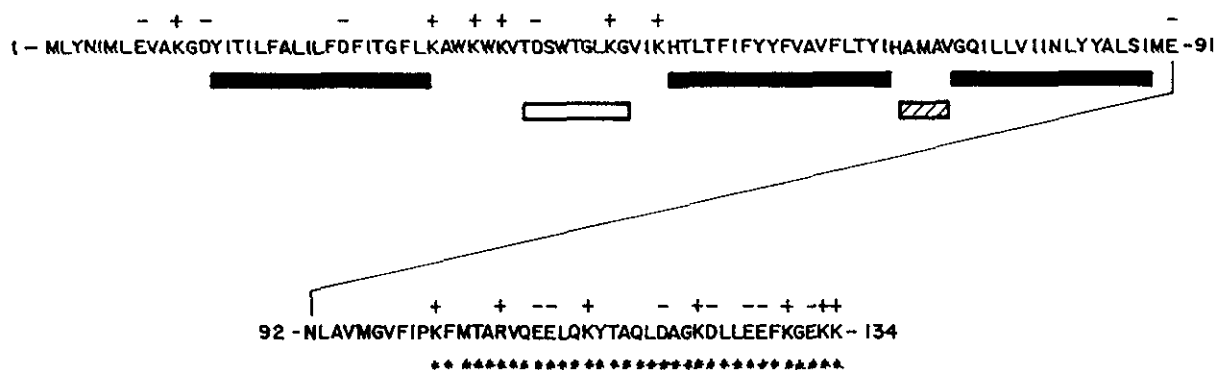
3.4.1. Clonación de la ORF 21 que codifica la holina y de la ORF22 que codifica la lisozima en *E. coli*

Los genes líticos son, en último término, los responsables de la lisis celular por lo que, para poderlos estudiar hay que regular negativamente su expresión. Con este fin, se utilizó el vector pNM185 para su clonación, un vector de expresión de amplio rango de

A



B



C

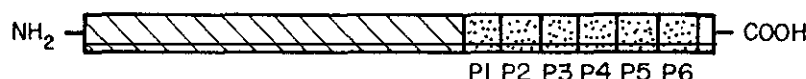


Figura 20. Holina y lisozima de Cp-1. A) Localización de la ORF21 y ORF22 que codifican la holina y la lisozima, respectivamente. PL8 y PL9, representados por puntas de flechas, indican los promotores en tándem que preceden al gen de la holina. B) Predicción de la estructura secundaria de la holina. En cajas negras se dibujan las presuntas hélices transmembrana, en caja blanca o rayada se indican los presuntos giros β de mayor o menor fiabilidad, respectivamente. En la parte superior de la secuencia se indican las cargas positivas (+) o negativas (-) de los aminoácidos. Los asteriscos indican el dominio C-terminal muy cargado. C) Organización modular de la lisozima. La región rayada representa el dominio N-terminal, las seis regiones punteadas, (P1-P6), representan motivos de unión a colina.

huésped, que sin embargo no se replica en neumococo. Este plásmido lleva el promotor Pm de la ruta *meta* del operón TOL, regulado positivamente. La transcripción a partir del promotor Pm es inducida por el producto del gen *xyl S* (Mermod y cols., 1986), después de su activación por moléculas efectoras como el 3-metilbenzoato (3-MB). El gen *cphI* y los dos genes líticos juntos, *cphI* y *cplI*, se amplificaron por la técnica de PCR como se describe en Materiales y Métodos, apartado 8, creándose los sitios de restricción *SacII* en el extremo 5' del DNA amplificado y *SacI* en el 3'. Como se indica en la Figura 21 los productos de la amplificación se clonaron en pNM185, construyéndose el pAMR11 que contiene el gen de la holina (*cphI*) y el pAMR12 con los genes *cphI* y *cplI*. Por otra parte, el gen *cplI* se había clonado previamente en el laboratorio en el vector de expresión pIN-III-A3 dando lugar al plásmido pCIP100 (Sanz y García, 1990).

3.4.2. Expresión inducible de los genes *cphI* y *cplI* en *E. coli*

Si el producto del gen *cphI* es una holina, su expresión debería inducir una lesión en la membrana citoplásmica y por tanto sería letal para la célula huésped. Después de la inducción del gen *cphI* con la adición de 3-MB se observó una inhibición del crecimiento del cultivo (Figura 22A) y su viabilidad se redujo 2,5 órdenes de magnitud (Figura 22B). Sin embargo, no se producía ningún efecto en el crecimiento de los cultivos que expresaban la lisozima (Figura 22A), ni se afectaba su viabilidad (Figura 22B). Por otra parte, cuando se expresaban conjuntamente los genes *cphI* y *cplI*, se observaba una disminución en la turbidez de los cultivos (Figura 22A) y una reducción de 2 órdenes de magnitud en su capacidad de formación de colonias. Estos datos indican que la holina de Cp-1 produce la muerte celular pero no lisa el cultivo y, por otra parte, que la lisozima sólo produce la lisis del cultivo cuando está coexpresada con la holina, lo que sugiere que la holina produce una lesión en la membrana permitiendo así el paso de la lisozima al periplasma.

3.4.3. El producto del gen *cphI* complementa la mutación *S* del fago lambda

Los resultados anteriores sugieren que la membrana es la diana del producto del gen *cphI*, como en el caso de la proteína *S* (holina) del fago lambda. Por otra parte, el fago lambda *Sam7* contiene una mutación ámbar en el gen *S* y es incapaz, por tanto, de inducir la

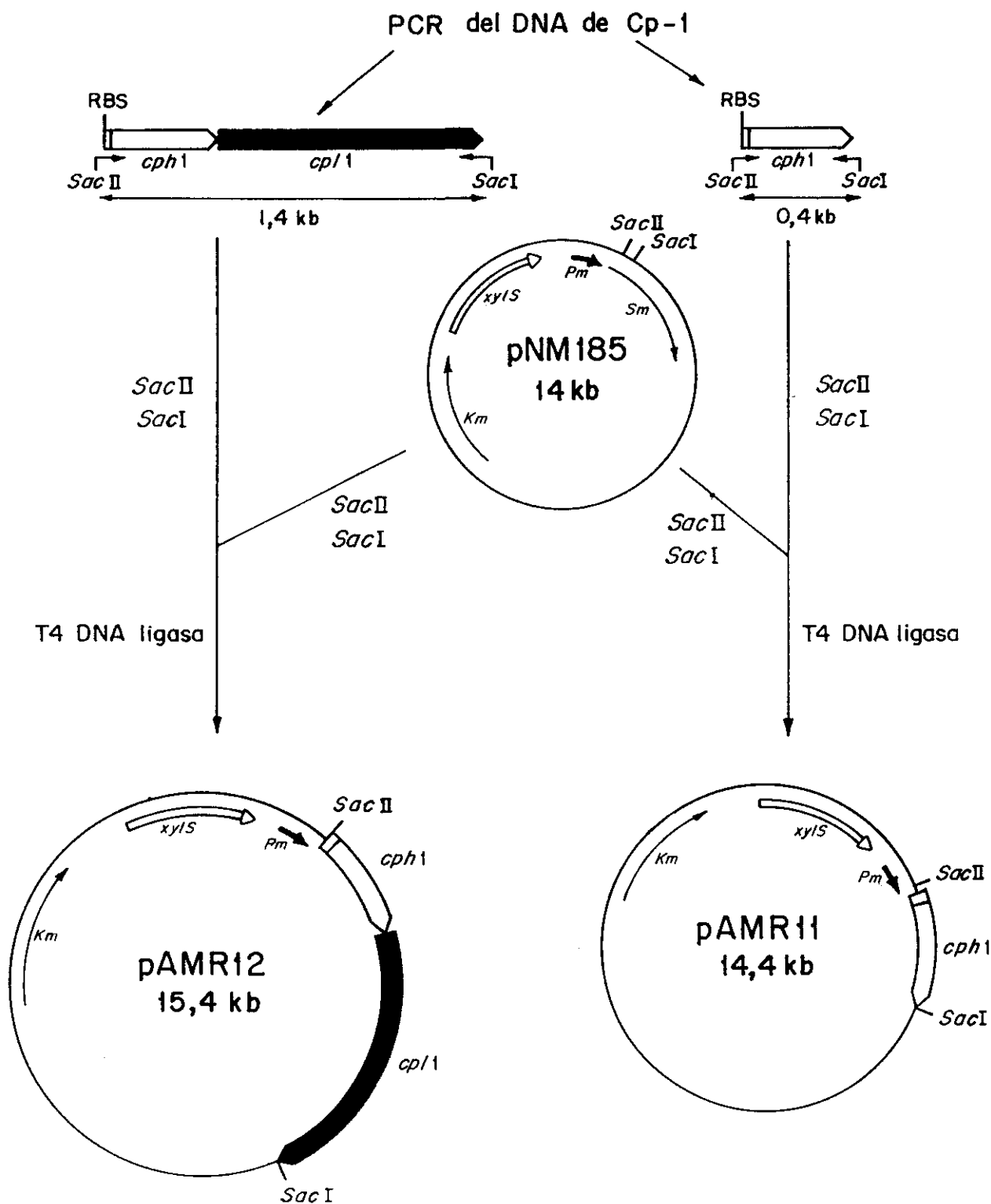


Figura 21. Construcción de pAMR11y pAMR12. Las flechas blancas gruesas representan el gen *cph1* y las flechas negras gruesas indican el gen *cpl1*. El rectángulo blanco que precede a *cph1* representa su RBS. Pm indica un promotor inducible de la ruta TOL.

lisis de células no supresoras de *E. coli*, puesto que la holina es esencial para la actividad de la lisina R de lambda (Reader y Siminovitch, 1971). Por tanto, para tratar de demostrar la función de la holina se hicieron pruebas de complementación utilizando este fago mutante en la infección de la estirpe HB101 de *E. coli*. Los resultados se muestran en la Tabla 14, donde se puede observar que el mutante del fago lambda es incapaz de inducir la lisis de HB101, huésped no supresor, pero produce placas en esta cepa cuando se expresa el gen *cphI*. Como se esperaba, este fago mutante no produce placas de lisis sobre esta bacteria cuando expresa el gen *cplI* (análogo al gen *R*). Puesto que el gen *cphI* complementa eficientemente el defecto del gen *S* del fago lambda se puede proponer que los productos de los genes *S* y *cphI* desempeñarían funciones similares.

Tabla 14. Complementación de la función del gen *Sam7* del fago lambda por el producto del gen *cphI*

Cepa de <i>E. coli</i>	Gen de interés	Lambda <i>Sam7</i> ^a	
		Con 3-MB ^b	Sin 3-MB
LE392 ^c	-----	+	+
HB101	-----	—	—
HB101(pCIP100)	<i>cplI</i>	—	—
HB101(pAMR11)	<i>cphI</i>	+	—
HB101(pAMR12)	<i>cplI</i> y <i>cphI</i>	+	—

Las placas de las cepas transformadas se incubaron a 30°C, el resto a 37°C. En los casos en los que sea preciso se añade el antibiótico correspondiente.

^a — y +, indican complementación negativa y positiva, respectivamente.

^b 3-MB se añadió a una concentración final de 0,2 mM.

^c Cepa supresora *supF58* de *E. coli* de la mutación ámbar de gen *S* del fago lambda.

3.4.4. Inducción de la actividad de LytA en *E. coli*

El gen que codifica la principal autolisina de *S. pneumoniae*, la amidasa LytA, se clonó y se expresó en *E. coli* sin causar efectos negativos visibles sobre el crecimiento y la viabilidad del cultivo (García y cols., 1986a). Asimismo, estudios sobre la localización celular de esta enzima en una cepa de *E. coli* que expresa constitutivamente el gen *lytA*

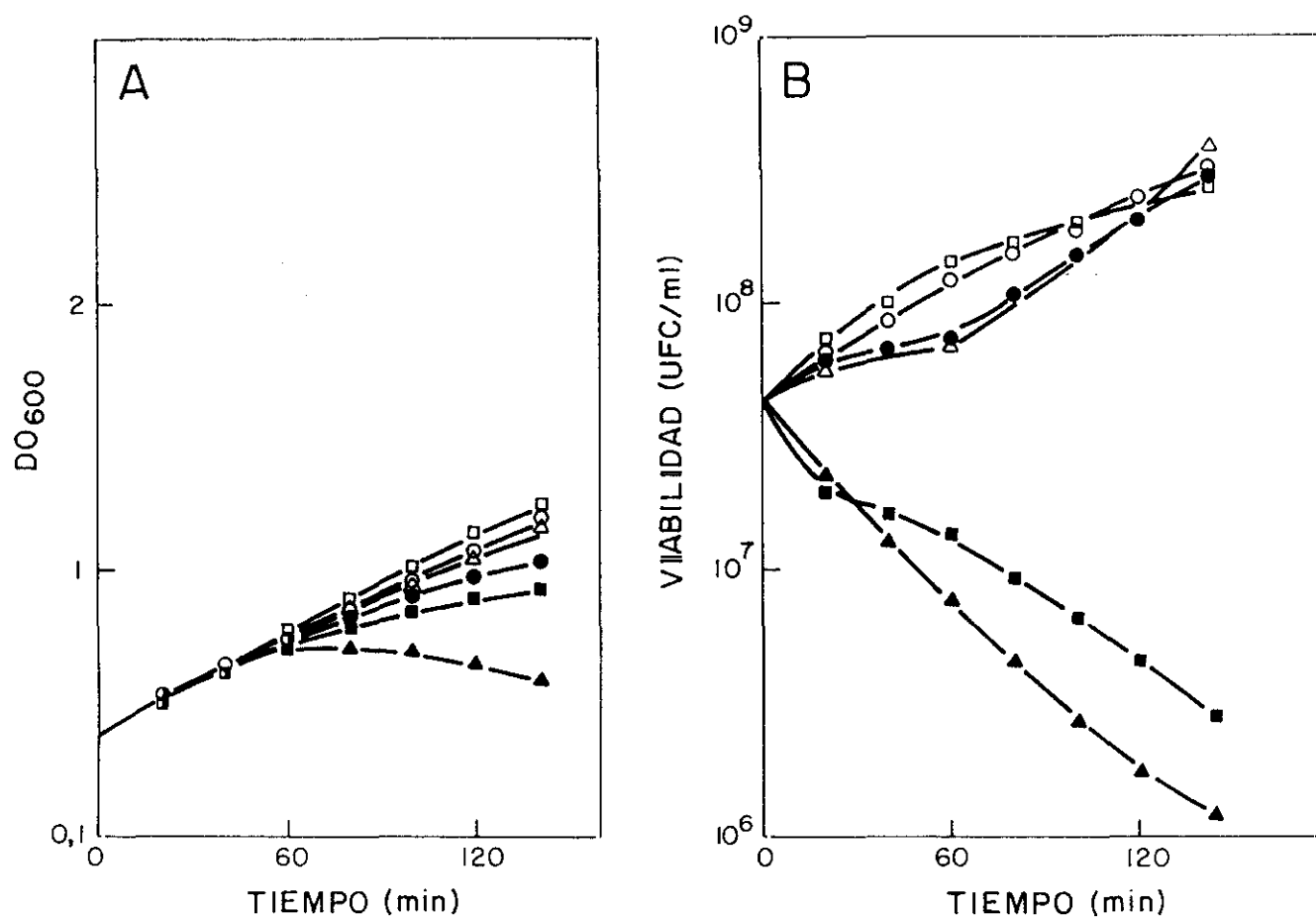


Figura 22. Expresión de *cph1* y *cpl1* en *E. coli*. A) Curvas de crecimiento. Cultivos de HB101(pNM185) (○, ●), HB101(pAMR11), (□, ■) y HB101(pAMR12) (△, ▲), sin inducir (símbolos blancos) o inducidos en el tiempo 0 con 3-MB (símbolos negros). Los cultivos se incubaron a 30°C. B) Viabilidad de los cultivos anteriores, expresada en UFC/ml.

indicaban que esta proteína está unida débilmente a la cara externa de la membrana citoplasmática, y se propuso que esta interacción con la membrana podría representar un mecanismo de regulación negativa para la actividad de esta enzima lítica (Díaz y cols., 1989). En este punto hay que recordar que tanto la amidasa LytA como la lisozima CplI son enzimas que dependen de colina para su actividad. Sin embargo, se ha comentado anteriormente que cuando la lisozima CplI se expresa junto a la holina CphI, los cultivos de *E. coli* se lisan al menos parcialmente. En este sentido se planteó estudiar si la lisis se producía también por efecto de LytA cuando se expresaba conjuntamente con la holina CphI. Para ello se transformó la cepa HB101(pAMR11), que lleva el gen *cphI* bajo el control del promotor inducible Pm, con pGL80 que lleva el gen *lytA* y lo expresa constitutivamente. El resultado, como se muestra en la Figura 23, demuestra que la lisis se produce cuando se añade el inductor 3-MB al cultivo. Este comportamiento concuerda también con lo descrito para la expresión conjunta de LytA y la holina del fago EJ-1 (Díaz y cols., 1996), lo que demuestra claramente que las enzimas líticas de neumococo no rompen la pared de *E. coli* si están confinadas en la membrana, pero son capaces de romper un número suficiente de enlaces de la pared celular, y provocar lisis, cuando el acceso a dicho sustrato está facilitado por la expresión de la holina.

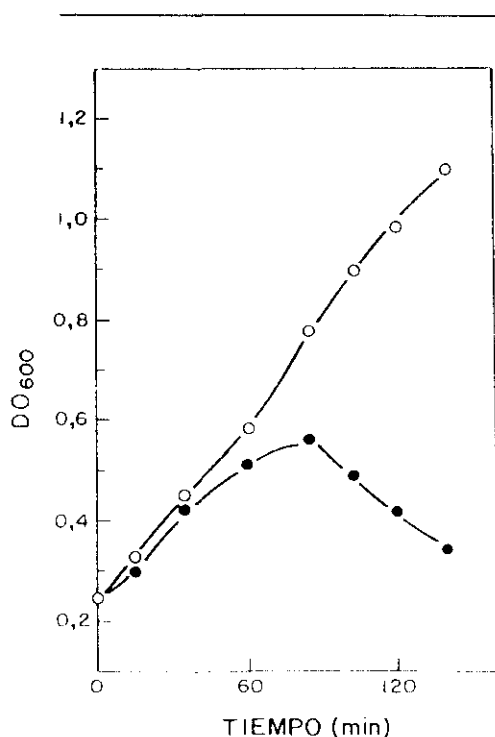


Figura 23. Expresión de *cphI* y *lytA* en *E. coli*. Cultivo de HB101(pAMR11)(pGL80) sin inducir (○) o inducido, en el tiempo 0, con 3-MB (●). La temperatura de incubación fue 30°C.

3.4.5. Clonación de los genes *cph1* y *cpl1* en neumococo

El objetivo de clonar estos genes en el sistema homólogo plantea serios inconvenientes, puesto que no existe ningún vector que replique en neumococo que posea promotores inducibles que permitan regular a voluntad la expresión de genes con efectos deletéreos, como son los implicados en el mecanismo de lisis fágica. Como posible alternativa se utilizó el vector pLSE1 y para evitar la posible letalidad de los productos de estos genes, éstos se clonaron sin su propio promotor. De esta manera, su expresión depende del promotor constitutivo de la proteína que proporciona resistencia a la Tc y que se encuentra separado aproximadamente a 1,6 kb del codón de comienzo del gen clonado. En la Figura 24 se esquematiza la estrategia de clonación. Con la mezcla de ligación se transformó la cepa C600 de *E. coli*, como paso previo para obtener el plásmido recombinante adecuado antes de transformar la cepa M31 de neumococo. La construcción correcta de los plásmidos recombinantes pAMR21 y pAMR22 se comprobó por cortes con enzimas de restricción, por hibridación y secuencia.

Para demostrar que se expresaban los genes clonados se hicieron experimentos de transcripción-traducción *in vitro* siguiendo el método descrito en el apartado 17.2 de Materiales y Métodos. Como se puede ver en la Figura 25, el plásmido pAMR21 codifica una proteína de masa molecular 15,5 kDa que coincide con la deducida de la secuencia de la holina (calle 3), por otra parte, el plásmido pAMR22 codifica dos proteínas de 15,5 y 41 kDa, que concuerdan con las masas moleculares deducidas de la secuencia de la holina y de la lisozima, respectivamente (calle 1). En la calle 2 aparecen las proteínas codificadas por el plásmido pLSE1 que sirve como control del experimento y de la especificidad de las proteínas codificadas por los correspondientes plásmidos recombinantes.

3.4.6. Expresión de los genes *cph1* y *cpl1* en neumococo

La expresión del gen de la holina de Cp-1 en neumococo produce, al igual que en *E. coli*, un retraso del crecimiento del cultivo que va acompañado de la pérdida de viabilidad del mismo (Figura 26). Sin embargo, aunque los efectos cualitativos en las dos bacterias son los mismos, en neumococo el efecto resultó ser menos letal, pues después de 4 horas de

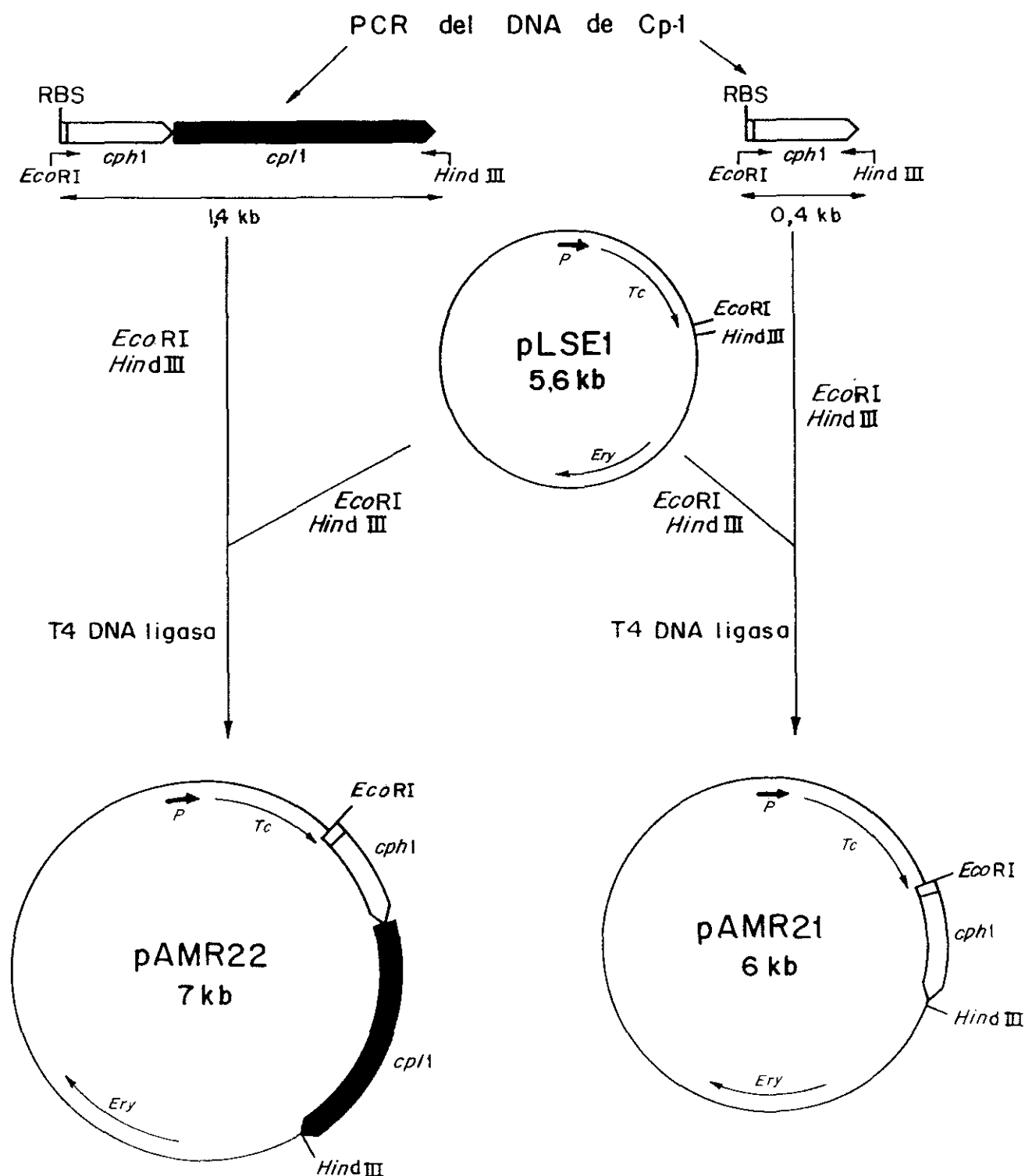


Figura 24. Construcción de pAMR21 y pAMR22. Las flechas blancas indican el gen *cph1* y las flechas negras el gen *cpl1*. El rectángulo blanco que precede a *cph1* representa su RBS. P indica el promotor del gen de resistencia a Tc.

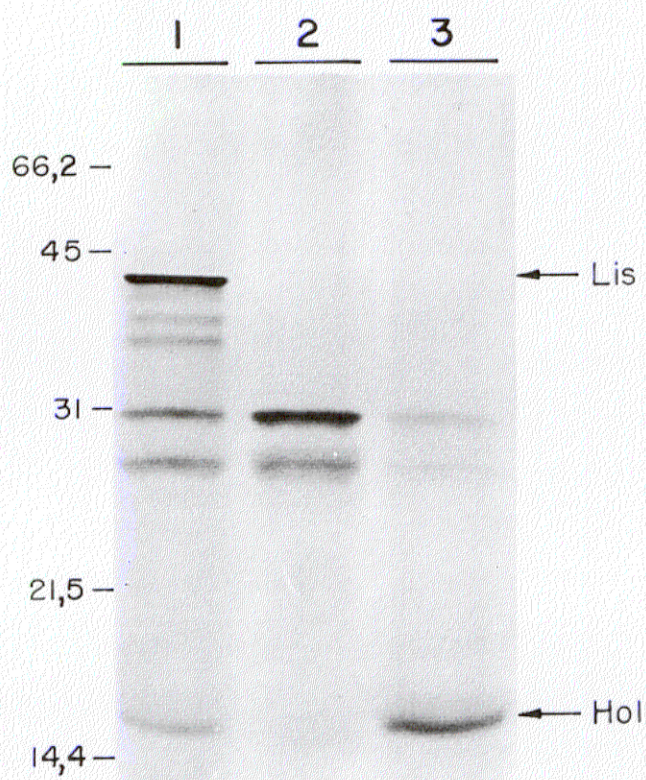


Figura 25. Transcripción-traducción *in vitro* de los plásmidos pLSE1, pAMR21 y pAMR22. Extractos conteniendo pAMR22 (calle 1), pLSE1 (calle 2) y pAMR21 (calle 3) se cargaron en un gel de poliacrilamida-SDS del 15%. Las flechas indican la posición de la holina (hol) y de la lisozima (lis). Los marcadores de peso molecular, en kDa, se muestran a la izquierda.

incubación el número de células viables disminuye media unidad logarítmica, mientras en *E. coli*, después de dos horas desciende 2 unidades logarítmicas.

Por otra parte, el efecto producido en una cepa *Lyt*⁻ de neumococo (M31) por la expresión de la lisozima Cpl1, ya había sido analizada anteriormente (Romero y cols., 1993), usando para ello una construcción similar a la presentada en esta Memoria (*cpl1* se clonó en pLSE1 desprovisto de su propio promotor), observándose que M31 se lisaba pero sólo dos horas después de alcanzar la fase estacionaria. Sin embargo, cuando se analizó el comportamiento de M31 transformado con pAMR22, que expresa los dos genes líticos (*cph1* y *cpl1*), se comprobó que la lisis se producía más rápidamente, sin que el cultivo llegase a alcanzar la fase estacionaria (Figura 26A). Estas observaciones sugerían claramente la implicación de la holina en la aceleración del proceso de lisis. No obstante, la pérdida de viabilidad de los cultivos de M31, antes de la lisis, era notablemente menor que en el caso paralelo de *E. coli* transformado con pAMR12.

Para corroborar que la holina ejercía el efecto fisiológico esperado, se intentó comprobar si moléculas de localización exclusivamente citoplasmática atravesaban la lesión o “agujero” formado por la holina en la membrana citoplasmática para ser detectadas en las estructuras periféricas. En primer lugar se evaluó el DNA, marcado radiactivamente, en los sobrenadantes de los cultivos M31(pLSE1), M31(pAMR21) y M31(pAMR22), con resultados negativos pues se encontró una mínima cantidad de radiactividad en los tres casos, antes de lisarse M31(pAMR22). A continuación se analizaron en M31 con y sin holina los niveles de neumolisina, la hemolisina de neumococo, que es una proteína exclusivamente citoplasmática (Johnson, 1977). Para ello se siguió la técnica de obtención de protoplastos descrita por Seto y Tomasz (1975) que permite la preparación de células desprovistas de pared y, por tanto, la detección de sustancias localizadas en el espacio periplásmico que han atravesado la membrana. De esta manera, se comprobó que M31(pLSE1) no liberaba hemolisina, pero en M31(pAMR21) y M31(pAMR22) sí se detectaban cantidades significativas de esta proteína. Ambos cultivos tenían niveles similares en el periplasma y era mucho más evidente al final de la fase exponencial de crecimiento (datos no mostrados). Estos resultados demostraban que la holina provocaba algún tipo de lesión en la membrana que permitiría el paso de determinadas proteínas y, previsiblemente, era responsable de la rápida translocación de la lisozima hacia su sustrato para desencadenar la lisis final del cultivo.

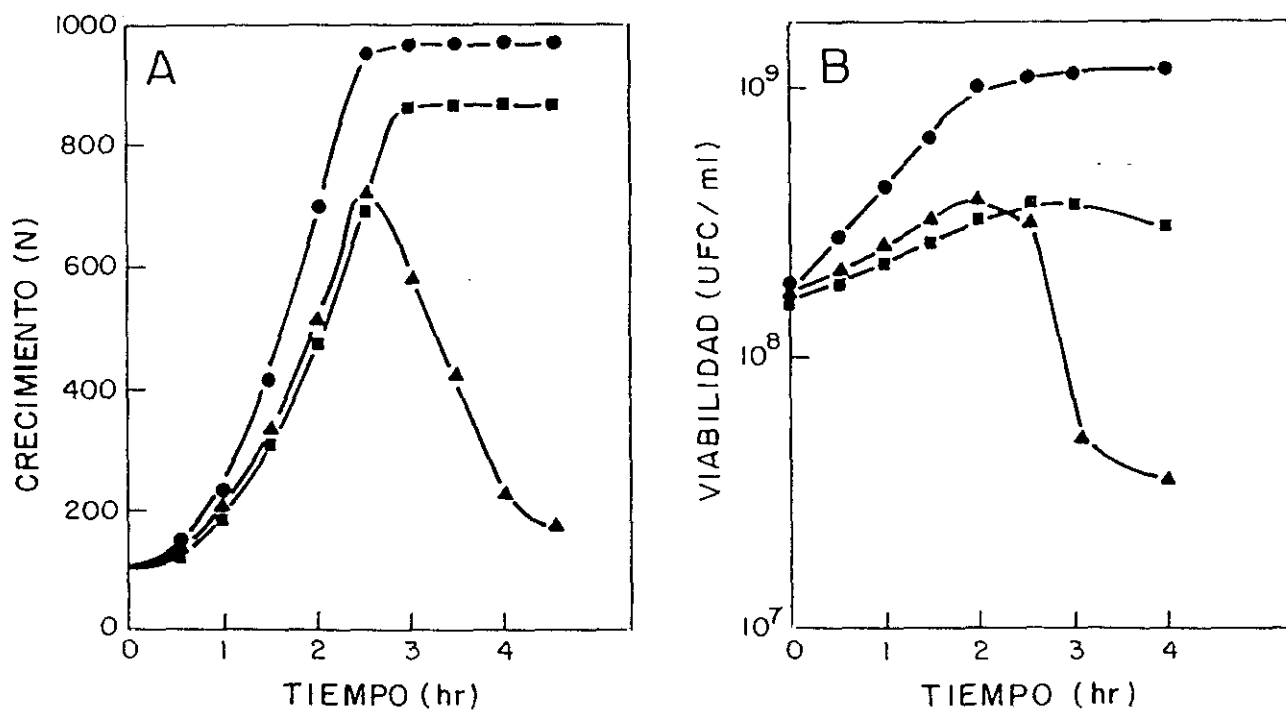


Figura 26. Expresión de *cph1* y *cpl1* en neumococo. A) Curvas de crecimiento. Cultivos de M31(pLSE1) (●), M31(pAMR21) (■) y M31(pAMR22) (▲), incubados a 37°C. B) Viabilidad de los cultivos anteriores, expresada en UFC/ml.

4. INICIACIÓN DE LA REPLICACIÓN DEL BACTERIOFAGO Cp-1

El fago Cp-1 replica su DNA mediante un mecanismo que utiliza TP como donadora del grupo 3'-OH libre que requiere la DNA polimerasa para iniciar el proceso. A esta conclusión se llegó cuando se incubaron extractos de neumococo infectados con [α - 32 P]dATP, dando como resultado la aparición de la TP marcada (García y cols., 1986). En los sistemas que replican por este mismo mecanismo se ha demostrado que los orígenes de replicación están en los extremos y, en algunos de ellos, se ha determinado el sitio exacto de iniciación (Méndez y cols., 1992; Caldentey y cols., 1993; King y van der Vliet, 1994). Basándonos en estos datos nos propusimos estudiar el sitio de iniciación de la replicación en el genoma del fago Cp-1.

Como paso previo, se comprobó que se podían utilizar extractos de neumococo infectado con Cp-1 como fuente de DNA polimerasa y TP y, por otra parte, oligonucleótidos sintéticos como molde. Del resultado positivo de este experimento se concluyó que no era necesario tener purificadas las proteínas implicadas en la iniciación de replicación para realizar este tipo de experimentos. Se hicieron ensayos de iniciación con las fracciones obtenidas en el proceso de eliminación de células enteras y restos celulares, llamadas sobrenadantes de alta velocidad, con las fracciones obtenidas después de la eliminación del DNA con polietilenimina, llamadas sobrenadantes de baja velocidad, y con las fracciones resultantes de la precipitación con sulfato amónico, tanto de células de neumococo infectadas por el fago como de las no infectadas, que sirven como control de especificidad (Figura 27). Como molde se usaron los oligonucleótidos sintéticos (12-mer) que llevan la secuencia del extremo derecho de ϕ 29 (Tabla 15). Se usaron estos oligonucleótidos como molde porque previamente se había demostrado para otros sistemas, como el fago ϕ 29, que se podía usar cualquier secuencia para obtener una respuesta positiva en la reacción de iniciación (Méndez y cols., 1992), aunque la eficiencia que se conseguía era variable. Además, en estos experimentos se comprobó que el Mn^{2+} era más eficiente como activador de la reacción que el ión Mg^{2+} . Como se puede ver en la Figura 27, la reacción de iniciación se debe a proteínas fágicas puesto que no hay formación de complejo de iniciación usando extractos no infectados, y depende además del DNA exógeno, ya que si no se añade este DNA no se produce el complejo TP-dAMP.

En todos los experimentos realizados se utilizó como control positivo del proceso y de la movilidad electroforética del complejo de iniciación TP-dNMP, la reacción de iniciación llevada a cabo con las proteínas DNA polimerasa y TP de $\phi 29$ purificadas. Dada la mayor masa molecular de la TP de $\phi 29$ (31 kDa) con respecto a la de Cp-1 (26,8 kDa), la movilidad del complejo de iniciación con proteínas purificadas de $\phi 29$ siempre era menor que la de Cp-1. Como control negativo se empleó una muestra en la que no se añadía DNA.

Tabla 15. Moldes de DNA de $\phi 29$ para los ensayos de iniciación y elongación

DNA	Secuencia (3'→5') ^a	dNTP preferido para la iniciación
$\phi 29$ oriR(12)	TTTCATCCCATG	dATP
$\phi 29$ T1C(12)	<u>CT</u> TTCATCCCATG	dATP
$\phi 29$ T2C(12)	TCTCATCCCATG	dATP
$\phi 29$ T3C(12)	TT <u>CC</u> ATCCCATG	dGTP
$\phi 29$ oriL(12)	TTTCATTTCGGGG	dATP

^a Los cambios producidos en la secuencia salvaje están subrayados. Las letras en negrita indican el sitio de iniciación en cada molde.

4.1. Formación del complejo de iniciación usando oligonucleótidos (29-mer) con la secuencia del extremo del genoma del fago Cp-1

Una vez comprobado que se podía trabajar con extractos crudos o semipurificados en lugar de con proteínas purificadas, se realizaron los experimentos de iniciación empleando oligonucleótidos que contenían la secuencia 3' de los extremos del DNA de Cp-1. Los extractos utilizados se sometieron a un paso de purificación utilizando para ello una columna de intercambio iónico, tal como se indica en el apartado 16.2 de Materiales y Métodos, con lo que se eliminaba una gran cantidad de proteínas del extracto. Como se muestra en la Figura 28, al incubar la fracción activa con el molde ori(29) (Tabla 16), [α -³²P]dATP y MnCl₂ como catión activador se produjo la formación del complejo de iniciación TP-dAMP. La eficiencia de esta reacción es proporcional a la cantidad de extracto soluble usado (calles 3-5). Extractos de neumococo no infectados no fueron

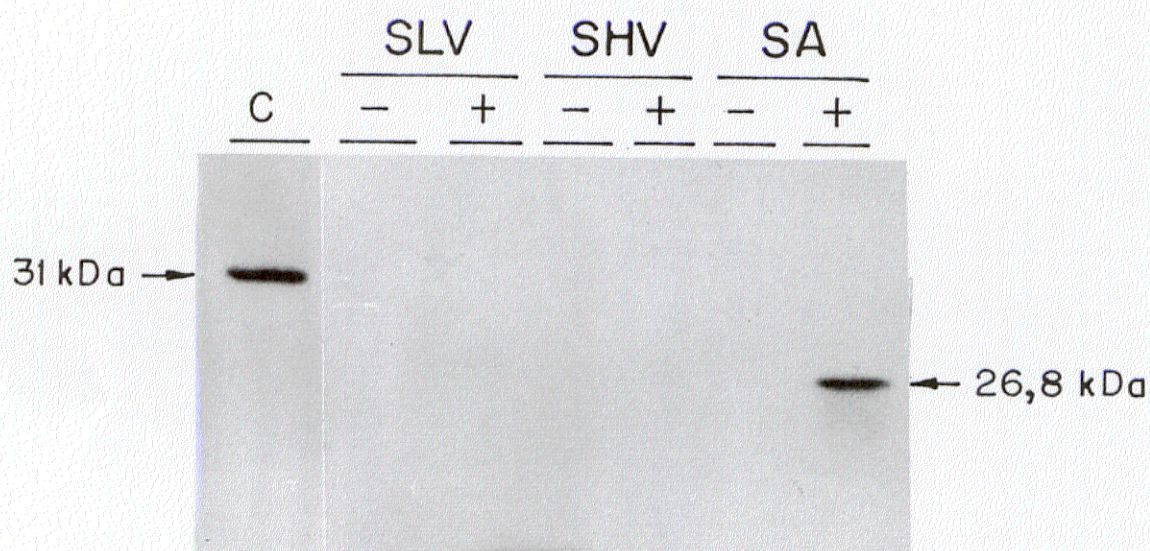


Figura 27. Ensayos de iniciación de la replicación con extractos crudos. Se usaron extractos de neumococo sin infectar (-) e infectados (+) y como molde oligonucleótidos (12 mer) con la secuencia del extremo derecho de $\phi 29$. SLV indica los sobrenadantes de baja velocidad obtenidos durante el proceso de obtención de extractos, SHV, sobrenadantes de alta velocidad y SA, la fracción obtenida después de precipitar las proteínas con sulfato amónico. C, control realizado con la DNA polimerasa y la TP purificadas de $\phi 29$. Las flechas indican las bandas radiactivas correspondientes a las TPs de $\phi 29$ (31 kDa) y de Cp-1 (26,8 kDa).

capaces de producir la formación del complejo de iniciación (calle 1) lo cual demostró que el complejo formado con los extractos infectados se debe exclusivamente a proteínas codificadas por el fago. Asimismo, se demostró que la formación del complejo fue estrictamente dependiente de la adición del DNA exógeno (calle 2) puesto que la formación no tiene lugar en ausencia de DNA. Además, para la iniciación de la replicación del DNA de Cp-1 no se requirió el origen completo (TP-dsDNA) pues es capaz también de utilizar eficientemente como molde moléculas de DNA de cadena sencilla, como ya se había descrito para otros sistemas con TP que replican por el mismo mecanismo (Méndez y cols., 1992; Caldentey y cols., 1993; King y van der Vliet y cols., 1994).

Tabla 16. Moldes de DNA de Cp-1 para las reacciones de iniciación y elongación

DNA	Secuencia (3'→5') ^a	dNTP preferido para la iniciación	Eficiencia de iniciación (%) ^b
ori(29)	TTTCGTACATGAGGGGGTGGGGAAAAAGT	dAMP	100
ori(29) T1C	<u>CT</u> TCGTACATGAGGGGGTGGGGAAAAAGT	dAMP	34
ori(29) T2C	T <u>CT</u> CGTACATGAGGGGGTGGGGAAAAAGT	dAMP	0,3
ori(29) T3C	TT <u>CC</u> GTACATGAGGGGGTGGGGAAAAAGT	dGTP	5,2
ori(29) T3A	TT <u>A</u> CGTACATGAGGGGGTGGGGAAAAAGT	dATP	<0,03
ori(29) T3G	TT <u>G</u> CGTACATGAGGGGGTGGGGAAAAAGT	dCTP	<0,03
ΔT1	<u>_</u> TTTCGTACATGAGGGGGTGGGGAAAAAGT	dGTP	0,03

^a Los cambios producidos en la secuencia salvaje están subrayados. Las letras en negrita indican el sitio de iniciación en cada molde.

^b La eficiencia de iniciación se expresa en relación con la actividad obtenida con ori(29).

4.2. La iniciación de la replicación del DNA de Cp-1 está dirigido por el tercer nucleótido

Como se ha dicho anteriormente, la formación *in vitro* de TP-dAMP es estrictamente dependiente de la presencia de DNA. La primera cuestión a resolver era si el DNA realmente actúa como molde para la formación del complejo de iniciación o simplemente estimula la reacción. Para distinguir entre ambas posibilidades, se ensayó la formación del complejo con cada uno de los cuatro nucleótidos marcados, en presencia del

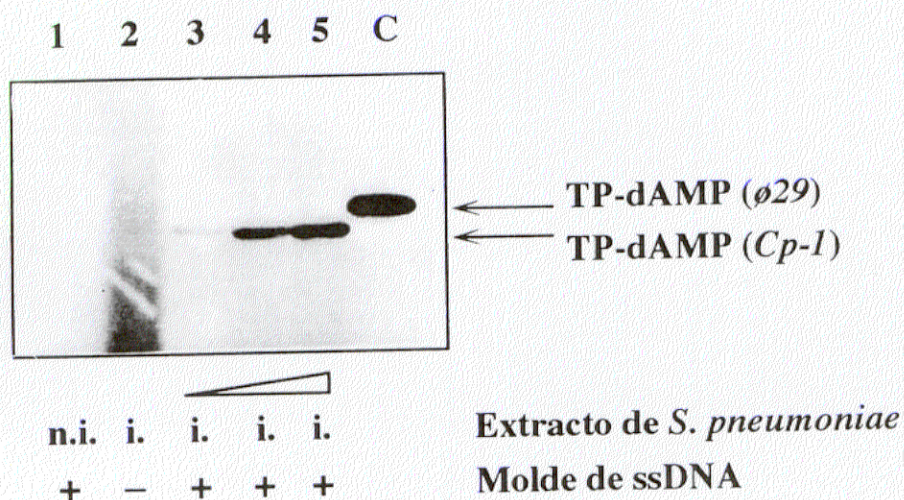


Figura 28. Formación *in vitro* del complejo de iniciación TP-dAMP. Las reacciones se realizaron usando como molde el oligonucleótido $\phi 29$ oriR(29). i., fracción parcialmente purificada de extractos de neumococo infectados con Cp-1. n.i., fracción de extractos de neumococo sin infectar. Las cantidades utilizadas en las reacciones fueron 0,5 μ l (calle 3), 1 μ l (calle 4) y 2 μ l (calle 1, 2 y 5). C, reacción de iniciación, en las mismas condiciones anteriores, pero usando DNA polimerasa y TP, purificadas de $\phi 29$.

oligonucleótido ori(29) que tiene la secuencia terminal 3' del DNA de Cp-1 (Tabla 16). Como se muestra en la Figura 29, el dAMP se unió eficientemente a la TP, mientras que no ocurrió lo mismo con el resto de dNTPs, lo que indica claramente que la formación del complejo TP-dAMP está dirigida por una T de la molécula del DNA.

El DNA de Cp-1 tiene una repetición de tres timinas (T) en sus extremos 3' (Figura 4), lo que sugiere que alguna de ellas tiene que ser la que dirija la reacción de iniciación. Para comprobarlo se sintetizaron oligonucleótidos con una sola sustitución en cada una de las tres timinas (ori(29) T1C, ori(29) T2C, ori(29) T3C) (Tabla 16) y se usaron como molde en las reacciones de iniciación con cada uno de los dNTPs marcados radiactivamente. La sustitución de la primera T por C (ori(29) T1C) no produjo un cambio significativo (Figura 29), el dNTP que se une preferentemente a la TP es el dAMP, y la eficiencia de incorporación del resto de los dNTPs fue en todos los casos menor al 1%. La eficiencia absoluta de incorporación de dAMP se redujo parcialmente con respecto a la secuencia original (34%) (Tabla 16). Igualmente, como también se puede ver en la Figura 29, la sustitución de la segunda T, no alteró la especificidad del dNTP que se va a unir a la TP, aunque este cambio produjo una drástica disminución de la eficiencia absoluta (0,3%) (Tabla 16). Por el contrario, la sustitución de la tercera T por C, no sólo produjo una reducción de la eficiencia absoluta (Tabla 16), sino que también provocó un cambio en la especificidad, puesto que es dGMP el que se unió a la TP y no dAMP como en los casos anteriores, (la reacción con dAMP fue sólo el 1% de la obtenida con dGMP). Este resultado indicó claramente que el inicio de la replicación de Cp-1 ocurre en el nucleótido complementario al que se encuentra en la tercera posición del extremo 3' del DNA. También indicó que el complejo TP-DNA polimerasa puede reconocer oligonucleótidos con la secuencia terminal alterada. Para corroborar estos resultados se sintetizaron oligonucleótidos con la tercera T sustituida por A (ori(29) T3A), por C (ori(29) T3C) y, por G (ori(29) T3G). La eficiencia de iniciación con estos nucleótidos disminuía drásticamente aunque la incorporación seguía siendo la del nucleótido complementario al mutado. Finalmente, el oligonucleótido $\Delta T1$ (Tabla 16) que es similar al ori(29) con la excepción de que en vez de tres timinas en el extremo 3' tiene dos, produjo la formación del complejo TP-dGMP (Figura 29), aunque fue extremadamente poco eficiente para la iniciación. Este resultado proporcionó un dato más para apoyar la idea de que es el tercer nucleótido el que dirige el proceso de iniciación de la replicación del DNA de Cp-1.

Molde	Secuencia terminal	Formación TP-dNMP				Eficiencia relativa, %			
		A	C	G	T	A	C	G	T
ori(29)	3' TTTCGT...					100	<0,1	0,1	<0,1
ori(29)T1C	3' CTTCGT...					100	<0,3	0,3	<0,3
ori(29)T2C	3' TCTCGT...					100	<10	15	<10
ori(29)T3C	3' TTCCGT...					1,3	<0,5	100	<0,5
ΔT1	3' TTCGT...					18	<10	100	<10

Figura 29. Especificidad del nucleótido de iniciación en el mecanismo de “protein-priming” de Cp-1. Las reacciones de iniciación se realizaron con cada uno de los cuatro dNTPs marcados. Se indica la secuencia 3' terminal de cada oligonucleótido usado como molde. En algunos casos, el tiempo de exposición que se muestra es 4 veces (ori(29) T2C; ori(29) T3C) y 30 veces (ΔT1) más largo que en el caso del ori(29) y el ori(29) T1C. Se indica para cada molde la eficiencia relativa de iniciación con cada uno de los cuatro posibles sustratos, considerando como 100% el nivel de incorporación del dNTP preferido.

4.3. Especificidad del origen de replicación: sitio de iniciación de extractos de Cp-1 sobre secuencias terminales del DNA de ϕ 29

La sustitución de la primera, segunda o tercera T de la secuencia molde de Cp-1, como se ha visto en el apartado anterior, reduce a diferentes grados la eficiencia de iniciación (Tabla 16). Esta reducción puede reflejar alguna especificidad de secuencia para el reconocimiento del origen de Cp-1 por el complejo TP-DNA polimerasa. Sin embargo, las mismas proteínas llevan a cabo la reacción de iniciación sobre oligonucleótidos (29-mer) cuya secuencia corresponde al extremo derecho 3' del genoma de ϕ 29 (Figura 30A), de la misma manera que se había comprobado con extractos crudos de células de neumococo infectadas por Cp-1. De hecho, la eficiencia de la iniciación *in vitro* usando oligonucleótidos (29-mer) con la secuencia de Cp-1 o con la de ϕ 29 fue aproximadamente la misma. Este resultado se puede explicar por la similitud existente entre los extremos del DNA de Cp-1 y de ϕ 29. Como se muestra en la Figura 30A, seis de los ocho primeros nucleótidos son idénticos en el origen del DNA de Cp-1 y en los orígenes derecho e izquierdo del DNA de ϕ 29. Hay que recordar que mientras los extremos del DNA de Cp-1 son iguales, el extremo derecho del DNA de ϕ 29 es distinto del izquierdo con la excepción de los seis primeros nucleótidos que forman las ITRs. El lado izquierdo del genoma de ϕ 29 es más similar a los extremos del de Cp-1 que el lado derecho aunque la reacción de iniciación con extractos de neumococo infectados por Cp-1 es más eficiente con los oligonucleótidos cuya secuencia corresponde al lado derecho, por razones difíciles de explicar. De los nucleótidos 11 al 23, 8 de los 13 residuos son iguales en el origen izquierdo del DNA de ϕ 29 y en los extremos del genoma de Cp-1.

El inicio de la replicación del genoma de ϕ 29 está dirigido por el segundo nucleótido del DNA molde (Méndez y cols., 1992), por lo que sería interesante estudiar cómo se comportan las proteínas de replicación de Cp-1 sobre las moléculas de DNA con los orígenes de replicación de ϕ 29, para determinar si el sitio preciso de iniciación está determinado por las características estructurales de las proteínas implicadas o por el DNA molde. Para ello se usaron como molde en la reacción de iniciación oligonucleótidos (29-mer) con una única variación en la secuencia del extremo derecho del DNA de ϕ 29 (Tabla 17). El resultado fue que la formación del complejo de iniciación TP-dAMP estaba dirigida por el tercer nucleótido del molde, tanto si el molde es de cadena sencilla como de cadena

doble (Figura 30B y Tabla 17) dado que el cambio de la tercera T por C dió como resultado la formación del complejo TP-dGMP en lugar de TP-dAMP. Como ocurre con los oligonucleótidos del DNA de Cp-1 mutados en la segunda y tercera T, también con los de $\phi 29$ disminuyó fuertemente la eficiencia absoluta de iniciación (Tabla 17). El hecho de obtener resultados similares con secuencias de Cp-1 y de $\phi 29$ sugiere que el sitio de iniciación está principalmente determinado por el complejo TP-DNA polimerasa más que por las señales internas impuestas por la secuencia del DNA.

Tabla 17. Moldes de dsDNA de $\phi 29$ para las reacciones de iniciación

DNA	Secuencia (3'→5') ^a	dNTP preferido para la iniciación	Eficiencia de iniciación (%) ^b
$\phi 29$ oriR(29)ds	TTTCATCCCATGTCGCTGTTGTATGTGGT	dATP	100
$\phi 29$ T1Cds	<u>CT</u> TTCATCCCATGTCGCTGTTGTATGTGGT	dATP	58
$\phi 29$ T2Cds	T <u>CT</u> CATCCCATGTCGCTGTTGTATGTGGT	dATP	0,5
$\phi 29$ T3Cds	TT <u>CC</u> ATCCCATGTCGCTGTTGTATGTGGT	dGTP	2,2

^a Los cambios producidos en la secuencia salvaje están subrayados. Las letras en negrita indican el sitio de iniciación en cada molde.

^b La eficiencia de iniciación se expresa en relación con la actividad obtenida con ori R(29)ds.

4.4. Replicación de los oligonucleótidos de ssDNA. Recuperación de la primera y segunda T

Los anteriores resultados demostraron que la replicación del Cp-1 comienza en la tercera posición complementaria al molde desde el extremo 3', lo que planteaba la incógnita de si las dos primeras posiciones se recuperaban o no en los primeros pasos de la replicación. Para contestar a esta pregunta se hicieron ensayos de elongación truncada con los mismos oligonucleótidos de Cp-1 utilizados para las reacciones de iniciación, es decir, tanto el que tiene la secuencia original como sus derivados mutantes.

Si se tiene en cuenta la secuencia de los extremos del DNA de Cp-1 (3'-TTTCGTACAT...), en presencia de ddTTP, la elongación sólo puede tener lugar hasta la primera A del molde (posición 7). Por tanto, si los dos nucleótidos terminales se recuperan

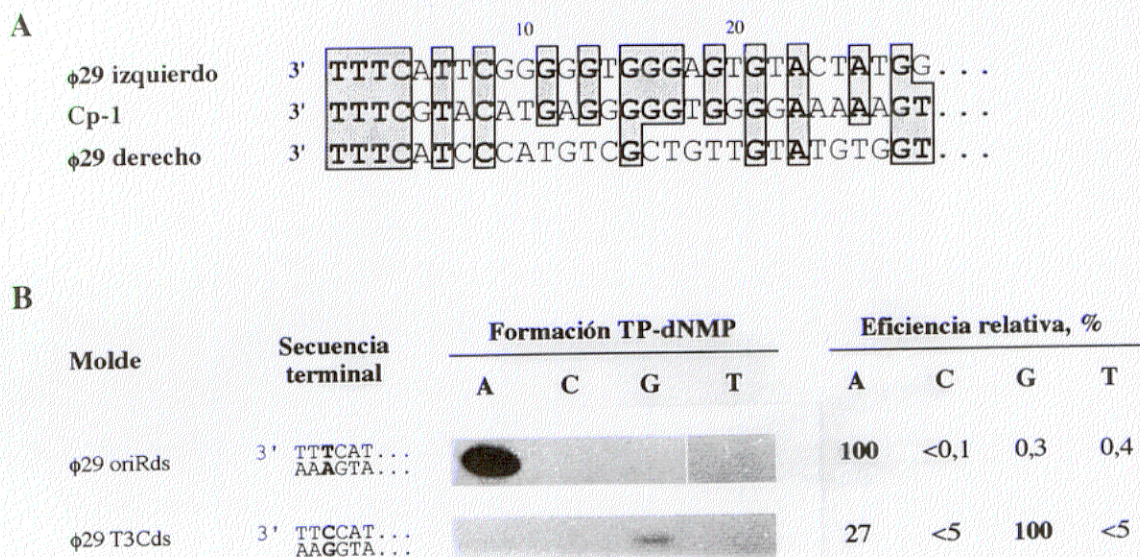


Figura 30. Replicación con DNA de ϕ 29 como molde. A) Comparación entre los orígenes de replicación del DNA de Cp-1 y de ϕ 29. Los residuos idénticos entre Cp-1 y los orígenes izquierdo y derecho de ϕ 29 están marcados con letras negras incluidas en cajas sombreadas. **B)** Sitio de iniciación con proteínas de Cp-1 sobre moldes de DNA de ϕ 29. Se usaron como molde oligonucleótidos de doble cadena con la secuencia del extremo derecho de ϕ 29 (ϕ 29 oriR(29)ds o la misma secuencia con la transversión T/C en el tercer par de bases (ϕ 29 T3Cds) para la formación del complejo TP-dAMP de Cp-1 con cada uno de los cuatro dNTPs. Se indica la eficiencia relativa de iniciación con cada uno de los cuatro sustratos, considerando como 100% el nivel de incorporación del dNTP preferido.

durante la elongación, la máxima longitud del producto de replicación sería TP-(dNMP)₇. Por el contrario, si ambos nucleótidos se pierden, el producto mayor de elongación sería TP-(dNMP)₅. Para poder diferenciar los diferentes productos de la elongación, TP-(dNMP)_n, se usaron geles de poliacrilamida-SDS de alta resolución (ver apartado 25.2 de Materiales y Métodos). Como se aprecia en la Figura 31, se observaron dos bandas con todos los oligonucleótidos probados. La banda de mayor movilidad electroforética corresponde al complejo de iniciación, TP-dAMP (TP-dGMP en el caso de ori(29) T3C) y, como era de esperar, su movilidad es independiente del molde usado (Figura 31). La banda superior corresponde al producto de elongación mayor, y su movilidad es claramente dependiente del oligonucleótido usado como molde (Figura 31). Puesto que no se ven otras bandas que se corresponderían con productos de elongación intermedios, es de suponer que sólo hay un producto de elongación final. Para poder asignar la longitud correspondiente a cada banda se asumió que el producto de elongación de la secuencia original (ori(29)) es TP-(dNMP)₇. En el caso de la molécula ori(29) T1C, la diferente movilidad sugiere que el producto es TP-(dNMP)₆, y en el caso del oligonucleótido ori(29) T3C, se correspondería con TP-(dNMP)₅. La eficiencia absoluta de replicación está limitada por la eficiencia de iniciación, por eso el oligonucleótido ori(29) T2C se replica muy mal. Sin embargo, la señal obtenida para la elongación de este molde, aunque débil se corresponde con la obtenida para la elongación de ori(29) (TP-(dNMP)₇). Para confirmar la correlación entre las bandas obtenidas y los productos de elongación, se hicieron ensayos de elongación truncada con los mismos oligonucleótidos pero esta vez en presencia de ddCTP, con el cual la elongación se detendría en la quinta posición desde el extremo 3' (Figura 31) y se formaría el producto TP-(dNMP)₅, en el caso de recuperarse los dos primeros nucleótidos. Cuando se analizaron los productos de elongación en presencia de ddTTP y de ddCTP en el mismo gel de electroforesis, el producto resultante de la elongación de la secuencia original en presencia de ddCTP se corresponde con una banda de movilidad electroforética similar a la obtenida para el producto del oligonucleótido ori(29) T3C con ddTTP, que sería la esperada para el producto TP-(dNMP)₅. Además el producto obtenido con ori(29) T1C parece ser una base más corta (TP-(dNMP)₄), y el obtenido con ori(29) T3C dos bases más cortas (TP-(dNMP)₃). La movilidad de esta última banda con respecto a la que se obtiene con el complejo de iniciación (TP-dGTP en este caso) se corresponde perfectamente con la diferencia de las dos bases entre ambos productos. Para concluir, el producto de elongación

obtenido con el molde ori(29) T2C en presencia de ddTTP tiene la misma movilidad que el producto obtenido del oligonucleótido con la secuencia original (TP-(dNMP)₅), aunque la señal es extremadamente débil.

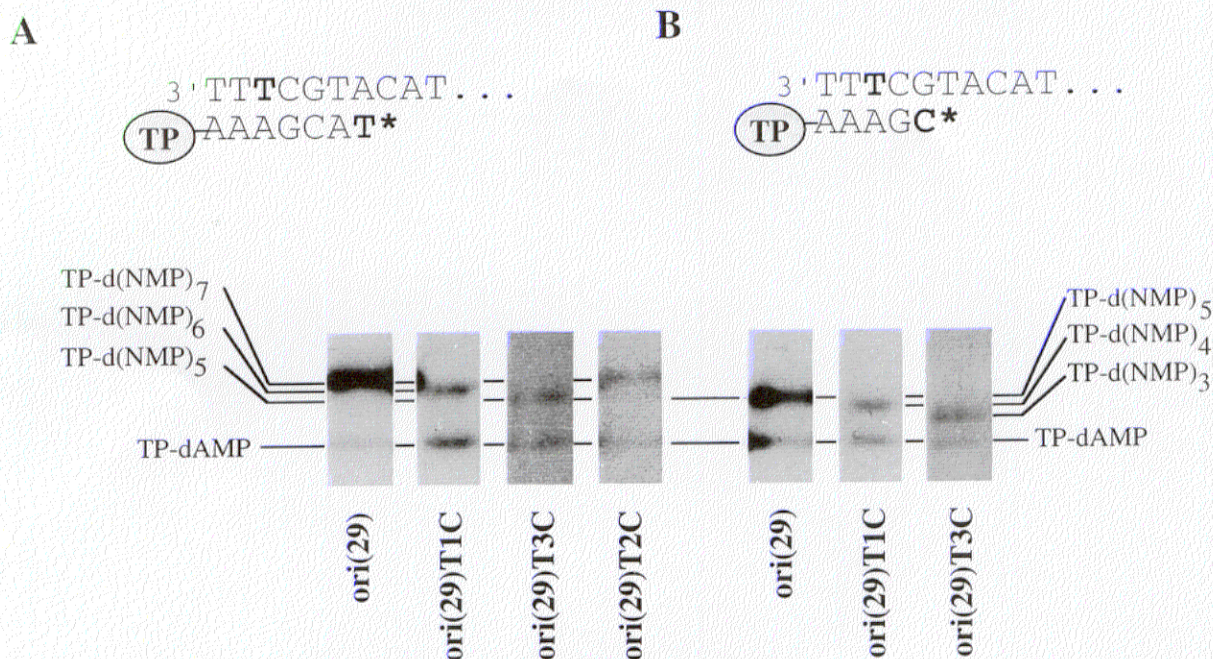


Figura 31. Elongación truncada sobre oligonucleótidos con la secuencia salvaje y mutada de Cp-1. A) Elongación en presencia de ddTTP. Se indica la molécula de DNA unida a la TP resultante de la replicación de ori(29) en presencia de ddTTP (T*), asumiendo que el primer y segundo nucleótido se recuperan durante la elongación. El sitio de iniciación en la molécula molde se señala en negrita. Las reacciones se realizaron y se analizaron como se indica en el apartado 25.2 de Materiales y Métodos. Debido a la baja eficiencia de replicación de ori(29)T2C, la calle correspondiente a este oligonucleótido se obtuvo por densitometría de la autorradiografía. **B)** Elongación en presencia de ddCTP. Método similar que en A, pero en presencia de ddCTP. Se indica el complejo de iniciación (TP-dAMP), y las correspondientes posiciones de los diferentes productos finales, TP-(dNMP)₃ a TP-(dNMP)₇ de cada ensayo de elongación truncada.

Estos resultados indican que durante el proceso de replicación de la secuencia original se recuperan las dos primeras posiciones del molde. Si no fuera así, los productos elongados en presencia de ddTTP serían iguales en todos los moldes utilizados, y también serían iguales los obtenidos en presencia de ddCTP, aunque dos bases más cortos que los de ddTTP. Además, es lógico que se recuperen porque, de otra manera, en cada ciclo de replicación se iría perdiendo información del genoma. La reiteración de las tres timinas parece crítica para esta recuperación, dado que sustituciones en la primera y en la tercera T del molde dan origen a productos de replicación intermedios.

En conjunto, estos resultados apuntan hacia un mecanismo de “sliding-back” en la transición entre la iniciación y la elongación del DNA, que se discutirá en el apartado 3 de Discusión, aunque no se podría descartar el mecanismo de “jumping-back”, es decir, el salto de la tercera a la primera posición del molde. El proceso de “sliding-back” se podría resumir de la siguiente manera (Figura 34A):

a) el heterodímero TP-DNA polimerasa reconoce el origen de replicación en forma de ssDNA. El centro activo de la DNA polimerasa posiblemente queda ubicado frente a la tercera base del extremo 3' del DNA.

b) la DNA polimerasa, dirigida por el tercer nucleótido, cataliza la adición de dAMP a la TP.

c) el complejo TP-dAMP realiza una translocación hacia atrás (“sliding-back”), de manera que el nucleótido unido a la TP se coloca frente a la segunda base del molde.

d) la DNA polimerasa utiliza de nuevo la tercera base del molde para extender el complejo TP-dAMP, dando lugar a TP-(dAMP)₂.

e) el complejo TP-(dAMP)₂ se translocaría de nuevo, de forma que el primer nucleótido unido a la TP se coloca frente al primer nucleótido del molde y el segundo frente al segundo.

f) la DNA polimerasa, por tercera vez, utiliza el tercer nucleótido del molde para extender el complejo TP-(dAMP)₂, formándose TP-(dAMP)₃. A continuación, la elongación ocurre normalmente.

IV. DISCUSIÓN

1. ORGANIZACIÓN FUNCIONAL DEL GENOMA DE Cp-1

Actualmente se conoce un número limitado de secuencias completas de genomas de bacteriófagos, sobre todo de aquellos que infectan bacterias Gram-positivas. Dentro de este grupo hay que citar el caso de $\phi 29$ y algunos fagos relacionados (Paces y cols., 1987) y más recientemente se ha determinado la secuencia de los fagos de *Lactococcus* bIL67 y c2 (Schouler y cols., 1994; Lubbers y cols., 1995) así como la del fago L5 de *Mycobacterium tuberculosis* (Hatfull y Sarkis, 1993). Por tanto, Cp-1 es el primer fago de *Streptococcus* cuyo genoma ha sido secuenciado completamente. Puesto que la genética de neumococo no está muy desarrollada debido, entre otras cosas, a que no se dispone de cepas supresoras ni de mutantes fágicos, el estudio de la secuencia de este genoma constituye una importante herramienta de trabajo para avanzar en los conocimientos de la genética y de la biología molecular de ese microorganismo.

Uno de los datos notables que se deducen de la secuencia de nucleótidos es la ausencia de dianas para algunas enzimas de restricción, tanto las que reconocen 6 pb (*EcoRI*, *BamHI*, *PstI*), como las que reconocen sólo 4 pb como *Sau3A*. Este hecho ha sido relacionado, en el caso del fago PRD1, con su amplio rango de huésped (Bamford y cols., 1991). No obstante, hasta ahora, se ha descrito que Cp-1 infecta solamente a neumococo y a *S. oralis* (Ronda y cols., 1989), otra bacteria filogenéticamente próxima y que contiene también colina en su pared celular, lo que podría explicar este aparente, limitado rango de huésped. Por otro lado, la información genética en el genoma del fago está muy compactada (93% es DNA codificante), pero existen varias regiones intergénicas de apreciable longitud, como entre las ORFs 2 y 3 (207 nt), 7 y 8 (224 nt), 8 y 9 (246 nt), 22 y 23 (1.111 nt) y b y c (284 nt). En la cadena complementaria de la mayor región intergénica (entre las ORFs 22 y 23) se localiza la ORF c; aún así, en la región del extremo derecho existe una región no codificante amplia entre los nt 17.448-17.899 (451 nt) y 18.139-18.559 (420 nt), lo cual contrasta con la idea de la gran compactación de la información genética en los genomas fágicos pequeños. Estos genomas suelen carecer de largas regiones intergénicas y, a veces, los fagos disponen de mecanismos moleculares que permiten que un solo gen codifique dos proteínas distintas, como el mecanismo de salto de fase, o genes que tienen otro codón de iniciación funcional en su interior, o los casos de reorganizaciones del genoma, como el de los fagos Mu, P1, P7 (Casjens y Hendrix, 1988). Otro ejemplo de aprovechamiento de la

información genética es el de PR4, bacteriófago que infecta *E. coli* y *Salmonella*, en el que un mismo gen da lugar a varias proteínas funcionales debido a un procesamiento postraduccional, ya que una pequeña proporción de la proteína mayoritaria de la cabeza se procesa para dar otras dos proteínas de la cápsida, con lo que, a partir de un gen, se originan tres proteínas funcionales (Myung y cols., 1994). En el caso de $\phi 29$, la región no codificante más larga se encuentra entre las ORFs 6 y 7 (aproximadamente 250 nt) pero, en esta región, se encuentran las zonas promotoras que controlan la transcripción temprana y tardía.

Si se analiza la secuencia de los extremos (comprendiendo las ITRs), se puede observar la existencia de tres posibles regiones promotoras consenso, dos en tándem y funcionales en los experimentos de “primer extension” (PE1 y PE2 en el extremo izquierdo, y PE4 y PE5 en el derecho) y una en dirección contraria, hacia los extremos del genoma, que no es funcional en experimentos de “primer extension” y cuyo significado se desconoce. Estos tres promotores recuerdan a la región de regulación de la transcripción temprana y tardía de $\phi 29$, donde A2_b y A2_c están en tándem y son los promotores de los mRNAs tempranos y A3 lo es de los tardíos en dirección contraria a los anteriores. Además, este hecho se podría corresponder con la posible inversión de esta agrupación de genes tempranos a lo largo de la evolución de estos fagos, con lo cual los tres promotores de los extremos de Cp-1 se corresponderían con los promotores A2_b, A2_c y A3 de $\phi 29$ (Figura 32B). El hecho de que estos tres promotores también existan en el extremo derecho del genoma de Cp-1, puede ser debido a que el origen de los mismos sea una duplicación repetida de una secuencia, seguida de una inversión, como sugiere el hecho de que las secuencias adyacentes a estos promotores sean similares en los tres casos (Figura 32A). Este dato indicaría que la inversión de los genes tempranos se produciría en el caso de $\phi 29$, porque si fuera al revés, sería más difícil tratar de explicar la existencia de estos tres promotores en ambos extremos.

Por comparación de secuencias de los productos de las ORFs localizadas en las regiones del genoma que transcriben genes tempranos con las existentes en las bases de datos, se pueden sugerir las funciones de las ORFs 4 y 5, como la TP y la DNA polimerasa, respectivamente. El resto de las ORFs que pertenecen a este grupo de genes tempranos, las ORFs 1, 2, 3, 6, 7, a, b y c, no muestran similitudes significativas con otras proteínas de función conocida, aunque sería lógico pensar que fueran proteínas implicadas en la

137

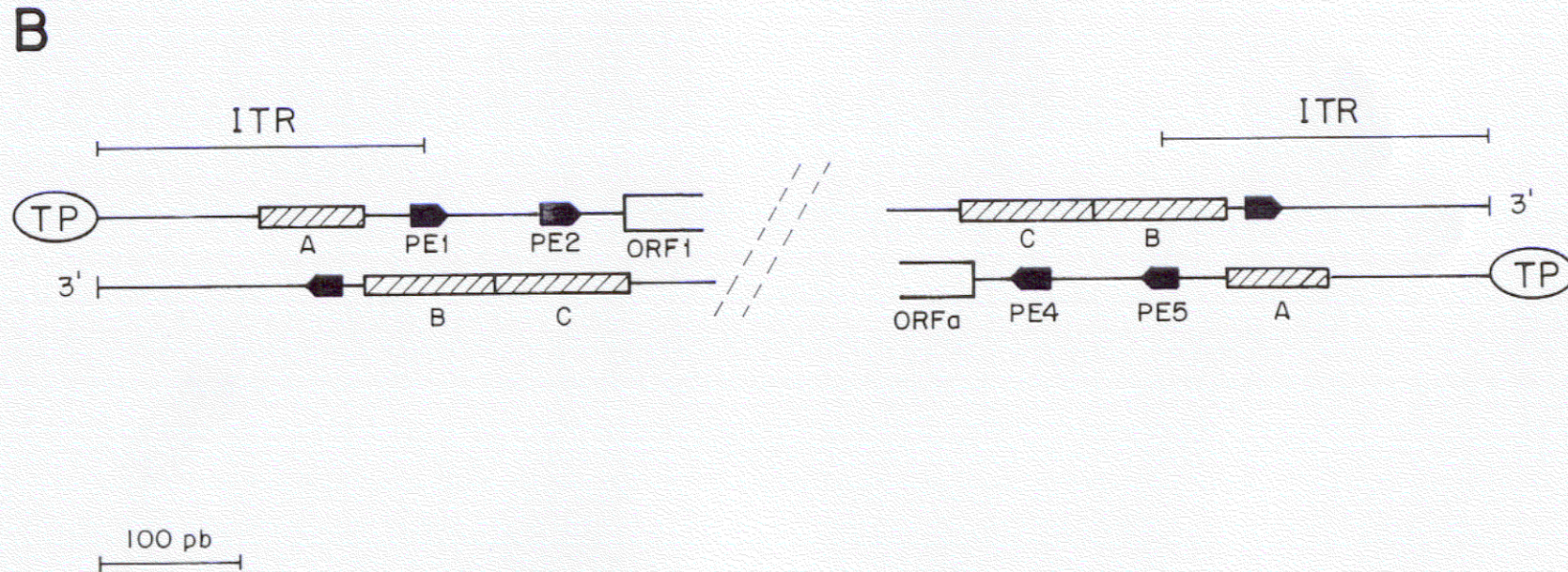
[illegible]

Figura 32. Secuencias de promotores en los extremos del DNA. A) Alineamiento de promotores en el extremo izquierdo del DNA del genoma. La secuencia del bloque A se indica en dirección 5'→3' y las de los bloques B y C corresponden a la cadena complementaria. En azul se representan los nt conservados en las tres repeticiones y en verde los que aparecen en dos de las tres repeticiones. Los números indican las mismas coordenadas de Cp-1 que en la **Figura 4**. B) Esquema de las repeticiones en los extremos del genoma. Las cajas rayadas representan las secuencias repetidas en cada extremo del genoma. Con flechas negras se indican los promotores, las cajas abiertas representan ORFs, y en los extremos 5' se esquematizan las TPs.

replicación o en la regulación de la transcripción como las proteínas p5, p6 y p4 del fago $\phi 29$, ya que este tipo de proteínas son las que se necesitan en los primeros estadios de la infección. A favor de esta hipótesis, está el hecho de que la ORF7 de Cp-1 posee un posible motivo de unión a DNA de cadena simple pero no totalmente conservado, así como un C-terminal ácido y un posible dominio de unión a ATP. Ambos aspectos son característicos de las proteínas SSB que, por el contrario, no suelen presentar homología a nivel de estructura primaria (Kim y cols., 1992).

Por otra parte, los genes tardíos se localizan agrupados en la parte central del genoma y se transcriben de izquierda a derecha como en el caso de $\phi 29$. En esta región se encuentran los genes que codifican las proteínas implicadas en el proceso de lisis, holina y lisozima, y otras ORFs que codifican proteínas estructurales como la de la cabeza (ORF9), el conector (ORF10), el collar (ORF11), u otras que intervienen en el ensamblaje, que, posteriormente, no se encuentran en el fago infectivo, como la proteína de andamiaje (ORF8) o la de encapsidación (ORF20). Estos datos concuerdan con los descritos en la mayoría de los fagos estudiados, donde las proteínas morfogenéticas se sintetizan coordinadamente en la etapa tardía de infección mientras las que controlan la expresión de los genes morfogenéticos son de síntesis temprana (Casjens y Hendrix, 1988). Además, los genes que codifican las proteínas estructurales que forman el virión no se localizan al azar dentro de los genomas de los fagos, sino que se encuentran agrupados. Así, según Stahl y Murray (1966) los genes cuyos productos interactúan directamente durante el ensamblaje y en el virión deben situarse muy cerca unos de otros, lo cual es estrictamente cierto en la mayoría de los fagos (Casjens y Hendrix, 1988).

El hecho de que algunos genes fágicos se transcriban desde diferentes promotores no es infrecuente. En el fago T7 podemos encontrar un ejemplo bien estudiado de promotores en tándem que transcriben los genes tempranos del extremo izquierdo de su genoma (Studier y cols., 1979). En el caso de $\phi 29$, los dos promotores en tándem, A2_b y A2_c, que inician la transcripción temprana, están separados por 100 pb y son utilizados, al menos *in vitro*, con diferente eficiencia (Murray y Rabinowitz, 1982), en contraste con los promotores en tándem que muestra Cp-1 (PE1-PE2, PE4-PE5, PL1-PL2 y PL8-PL9), en los que ambos se utilizan con la misma eficiencia. En el fago A511 de *Listeria monocytogenes*, se ha demostrado también la existencia de dos promotores en tándem precediendo a la ORF3 que se encuentra en el extremo 3' del gen que codifica la proteína

mayoritaria de la cabeza (Loessner y Scherer, 1995). Recientemente, en el fago Tuc2009 se han propuesto diferentes promotores precediendo al gen que codifica la presunta holina (Arendt y cols., 1994) al igual que en el fago $\phi 105$ de *B. subtilis* (Leung y Errington, 1995).

Con respecto a las secuencias de promotores, hay que resaltar que son pocos los genes de neumococo cuyos elementos de transcripción hayan sido bien caracterizados. Un análisis detallado de los promotores de neumococo caracterizados como tales hasta 1990 (Morrison y Jaurin, 1990) puso de manifiesto que aquellos son muy similares a los consenso de *E. coli*. Idéntica conclusión se obtuvo del estudio de los promotores tempranos de los fagos de *Bacillus* (Murray y Rabinowitz, 1982). El alineamiento de todos los promotores funcionales encontrados en el genoma de Cp-1 (Tabla 11) permite deducir una secuencia consenso que coincide también con las secuencias promotoras consenso de neumococo y de los fagos de *Bacillus* (Tabla 11). Sin embargo, la secuencia de los promotores de los genes tardíos difiere sustancialmente de los anteriores, aunque no se conoce si esto tiene alguna implicación en la regulación de los diferentes mRNAs.

Una de las peculiaridades que comparten algunos fagos y virus animales es el papel que juega en el proceso de empaquetamiento de DNA un RNA pequeño, llamado RNA piloto (pRNA). En el caso de $\phi 29$, el pRNA no es un componente del virión pero está presente en las precabezas durante el proceso de encapsidación, tanto en experimentos realizados *in vitro* como *in vivo* ((Guo y cols., 1987a, b; Anderson y Bodley, 1990). Su tamaño es de 174 nt y se transcribe a partir de un promotor localizado en la región de la izquierda del genoma (Guo y cols., 1987b). Sin embargo, el pRNA aislado es de 120 nt pues se hidroliza cuando en el procedimiento de obtención no se utilizan inhibidores de RNAsas (Wichitwechkarn y cols., 1989). Su estructura secundaria muestra dos dominios, uno más grande (dominio I) que contiene cuatro hélices y otro menor con dos hélices (dominio II) separados por 13 nt desapareados (Bailey y cols., 1990; Anderson y Bodley, 1990). La asignación funcional detallada de los diferentes dominios y hélices se está llevando a cabo por el grupo de Anderson, fundamentalmente mediante estudios de mutantes truncados y de mutagénesis dirigida (Grimes y Anderson, 1989; revisado por Anderson y Reilly, 1993; Reid y cols., 1994a, b). RNAs similares se han encontrado en las precabezas de fagos relacionados con $\phi 29$ y aunque sus estructuras primarias son muy diferentes, las secundarias están altamente conservadas (Bailey y cols., 1990). Asimismo, se ha demostrado la implicación del RNA en el empaquetamiento del DNA en el fago híbrido

lambda/ ϕ 29 (Donate y Carrascosa, 1991), en el del fago D108, relacionado con Mu (Burns y cols., 1990) y se ha propuesto su participación en poxvirus (Parsons y Pickup, 1990) y en adenovirus (Hatfield y Hearing, 1994) cuyo genoma replica valiéndose de un mecanismo similar al ϕ 29. Los resultados obtenidos en el sistema de Cp-1 sugieren también la existencia de una molécula análoga de pRNA, si bien el gen responsable de su síntesis se localiza en el extremo derecho del genoma y se transcribe hacia el extremo a partir del nt 18.569. El hecho de que el pRNA en ϕ 29 se empiece a transcribir a una distancia de 320 nt del extremo y en cambio el pRNA de Cp-1 a 775 nt puede ser debido a las hipotéticas duplicaciones de los extremos del genoma antes comentadas. Por otra parte, se ha sugerido que estos RNAs están implicados en la polaridad de la entrada del DNA en el proceso de su empaquetamiento a través de la complementaridad de bases existente entre el RNA y el extremo del DNA (Bjornsti y cols., 1983; Grimes y Anderson, 1989). No obstante, todavía no se conoce si la disposición de estos RNAs en posiciones opuestas del genoma en ϕ 29 y Cp-1 tiene relación con el extremo del DNA que se empaqueta primero, pero es interesante resaltar que el inicio de este RNA en el caso del Cp-1 está más alejado del extremo que en el caso de ϕ 29, de forma que no se localiza en la zona de ITRs y se pueden distinguir ambos extremos. Por tanto, la similitud entre las estructuras secundarias de estos pRNAs de fagos que infectan dos huéspedes diferentes, añade un nuevo interés filogenético al estudio de este RNA pequeño que juega un papel fundamental en el empaquetamiento del DNA del fago. En relación con esta prioridad en el empaquetamiento, en el sistema híbrido lambda/ ϕ 29 se ha observado que el conector de ϕ 29 une preferentemente el pRNA de dicho fago a otros RNAs de pequeño tamaño, lo que indica que se requiere un RNA específico para seleccionar el DNA apropiado para empaquetar (Valpuesta y cols., 1993). Además, la implicación de estos pRNAs en la encapsidación de ciertos genomas virales podría convertirles en una diana potencial de interés para el diseño de drogas antivirales.

Por otra parte, Cp-1 y ϕ 29 muestran similitudes morfológicas muy notorias y comparten el mismo mecanismo de replicación. Además, como ya se ha señalado, la organización de ambos fagos es similar y presentan proteínas muy similares. Estos datos sugieren que al menos una parte de los genomas respectivos puede haber divergido a partir de un ancestro común muy alejado en el tiempo. No obstante, el análisis global de la organización funcional revela que la principal diferencia entre los genomas de Cp-1 y ϕ 29 es la posición de los mRNAs tempranos de la parte izquierda del genoma, que corresponde a

los genes que codifican, entre otros, la TP y la DNA polimerasa. En $\phi 29$, este mRNA temprano se transcribe de derecha a izquierda, y la proteína temprana, p4, está implicada en la terminación de la transcripción temprana y la activación de la tardía que tiene lugar de izquierda a derecha (Salas y Rojo, 1993). Por el contrario, Cp-1 transcribe casi exclusivamente una sola cadena de DNA, de izquierda a derecha y no está claro que la transición de la transcripción temprana a la tardía esté controlada por una única región del DNA (Figura 33). En este sentido, la organización transcripcional de PRD1 (Bamford y cols. 1991) se parece más a la de Cp-1, ya que los genes tardíos se transcriben de izquierda a derecha y se localizan en el centro del genoma y los tempranos se encuentran en los extremos, transcribiéndose de manera similar a los de Cp-1 (Figura 33). Si se supone que los genomas de los fagos Cp-1 y $\phi 29$ evolucionaron a partir de elementos genéticos funcionalmente similares se puede especular con que la región de la izquierda de alguno de los genomas haya sufrido una inversión en algún momento durante dicha evolución. Por lo general, se ha observado una organización funcional similar y una evolución a partir de un ancestro común entre genomas o agrupaciones de genes entre bacteriófagos relacionados que infectan a la misma bacteria (Vasala y cols., 1993), en contraposición al caso de Cp-1 y $\phi 29$ que infectan a bacterias muy distintas entre sí. No obstante, se acepta cada día más la idea de que genomas de fagos no relacionados deben haber intercambiado parte de sus genes, como se ejemplifica muy bien en el caso de los genes de las fibras de la cola de los fagos lambda, Mu, P1, P2 y T4 (Sandmeier, 1994). Como resultado, genes individuales aparecen como mosaicos integrados por partes derivadas de conjuntos de genes comunes. Por tanto, la transferencia horizontal constituye el factor fundamental en la evolución de una parte específica de los genomas fágicos. Recientemente, Ackermann y cols. (1995) han postulado un origen común para los fagos con cola, en relación a una serie de propiedades comunes como la estructura de la cabeza y la cola, el genoma con DNA de doble cadena, grupos de genes con funciones relacionadas y la maduración de las proteínas de las precabezas por proteólisis. Además, la adquisición y reordenamiento de segmentos de genes bajo una fuerte presión selectiva es un hecho frecuente en bacterias aunque todavía no se haya estudiado extensamente. En este sentido nuestro laboratorio ha investigado la interrelación fago-bacteria desde el punto de vista de los genes de las enzimas líticas, sobre todo en el caso del fago atemperado HB-746 de neumococo (Romero y cols. 1990b; López y cols., 1995). Dicho sistema ha proporcionado datos experimentales que sugieren que

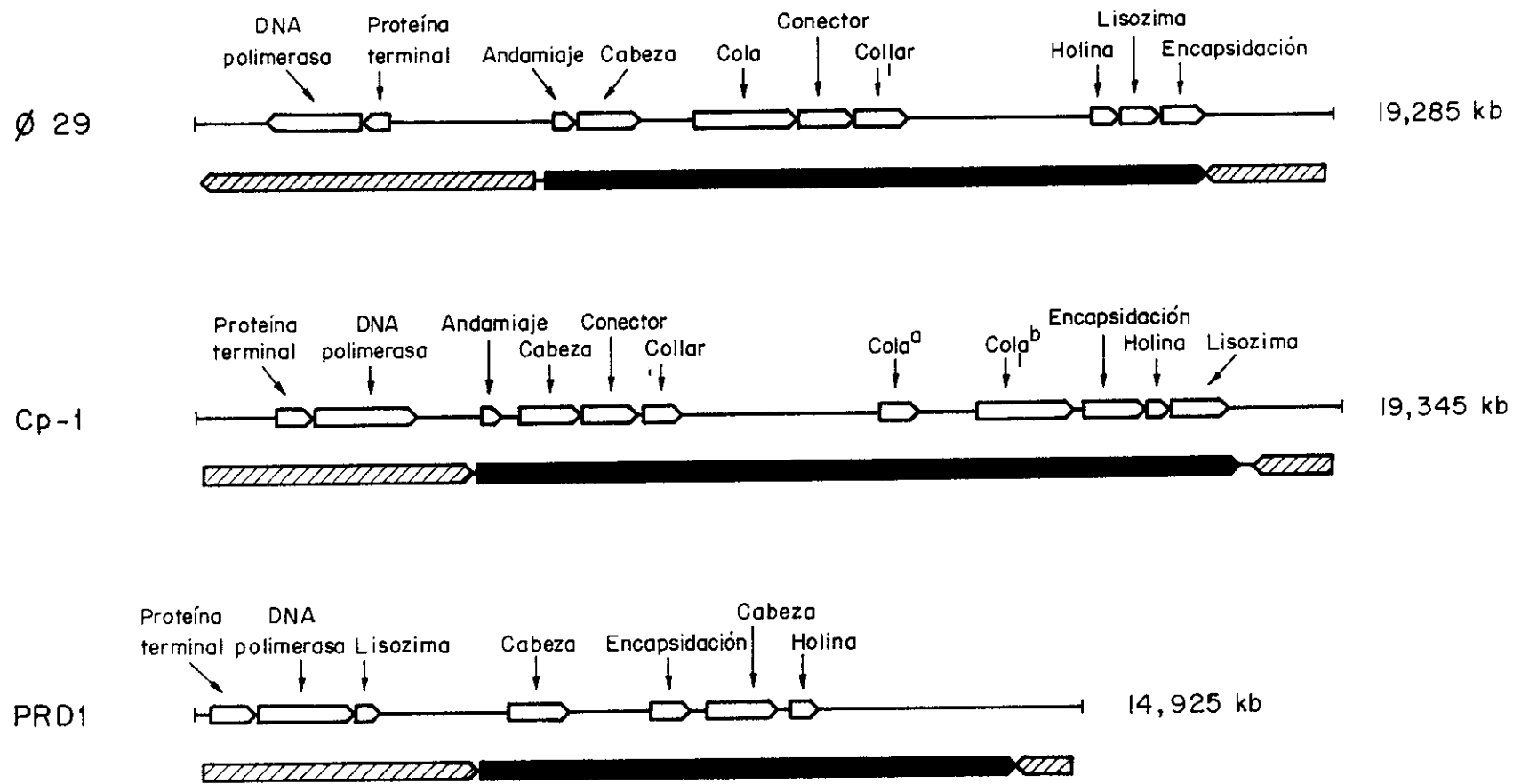


Figura 33. Organización funcional de los genomas de los fagos $\phi 29$, Cp-1 y PRD1. Con flechas blancas se indican algunas de las ORFs identificadas, con flechas rayadas se muestran los mRNAs tempranos y con flechas negras se representan los mRNAs tardíos de dichos genomas.

evolutivo, han dado lugar a enzimas líticas capaces de degradar la pared bacteriana. Se podría por tanto especular con que, mediante mecanismos de recombinación, el gen ancestral de la enzima lítica de Cp-1 adquirió el dominio de unión a colina que determinó su especificidad para reconocer y degradar paredes conteniendo colina lo que, por consiguiente, condicionan el estrecho rango de huésped que presenta el fago Cp-1. Por el contrario, la adquisición de este dominio no se habría dado en el caso de $\phi 29$ y, posiblemente, dicho evento habría determinado el punto de divergencia entre los dos fagos.

2. ANÁLISIS DE PROTEÍNAS CODIFICADAS POR Cp-1

Ya se ha comentado anteriormente la gran similitud entre distintas proteínas de $\phi 29$ y de Cp-1. Entre todas ellas, la DNA polimerasa de Cp-1 es la que posee un mayor grado de identidad con la correspondiente de $\phi 29$, particularmente en los motivos que se han demostrado fundamentales en el grupo B de las DNA polimerasas; por ejemplo, los motivos ExoI, ExoII y ExoIII, que son los responsables de la actividad exonucleasa 3'→5' y la capacidad de desplazamiento de banda (Soengas y cols., 1992), y los motivos 1, 2a, 2b, 3 y 4 que son los responsables de la iniciación de la replicación por el mecanismo de "protein-priming" y de las reacciones de polimerización del DNA. Además, la DNA polimerasa de Cp-1 también contiene una inserción, entre las regiones 1 y 2a, que es una característica compartida con las DNA polimerasas que replican mediante "protein-priming" (Braithwaite e Ito, 1993). Esta última región se ha sugerido que está implicada en la alta procesividad de la polimerasa de $\phi 29$ (Salas, 1991). Desde un punto de vista evolutivo, está claro que la DNA polimerasa de Cp-1 muestra unas regiones conservadas más relacionadas con el grupo de DNA polimerasas que comparten el mecanismo de "protein-priming" que con la única DNA polimerasa descrita hasta ahora en neumococo (López y cols., 1989). En este sentido la DNA polimerasa de *S. pneumoniae* contiene los dominios implicados en la polimerización y en la actividad exonucleolítica 5'→3'; sin embargo, en contraste con la mayoría de las DNA polimerasas, esta enzima carece de la actividad exonucleolítica 3'→5', a pesar de la identificación de algunos aminoácidos importantes característicos de los motivos de este dominio exonucleolítico (Díaz y cols., 1992b).

Por otra parte, la TP es otra de las proteínas implicadas en este mecanismo peculiar de replicación del DNA. La secuencia primaria de las proteínas terminales conocidas hasta

Por otra parte, la TP es otra de las proteínas implicadas en este mecanismo peculiar de replicación del DNA. La secuencia primaria de las proteínas terminales conocidas hasta hoy no muestra una gran similitud entre ellas, pero comparten perfiles hidropáticos similares, como ocurre con la proteína que codifica la ORF4 de Cp-1. Además, la localización de la ORF4 en el genoma fágico, precediendo al gen que codifica la DNA polimerasa es la misma que la que presentan los genes de la TP de los fagos PRD1, ϕ 29, PZA y GA1. En los casos en que se ha determinado el aminoácido concreto de unión al DNA, se ha visto que éste está situado en una zona de giro β precedido de una región de hélice α . En el caso de la TP de Cp-1, se ha demostrado que se trata de una thr (García y cols., 1986b), aunque la asignación precisa del residuo concreto, basado en entornos de estructura secundaria como se ha descrito en otros sistemas, no es posible por el momento.

En lo que se refiere a las proteínas de la cola, Cp-1 sintetiza dos proteínas que se parecen a las regiones N- y C-terminales de la única proteína de la cola de ϕ 29. Ésta cumple dos funciones diferentes, por una parte estructural de formación de la cola por donde el fago inyecta su DNA en el huésped y, por otra, la de estabilizar el empaquetamiento del DNA en la precabeza (Lee y Guo, 1994). Por otra parte, en el fago c2, se ha demostrado la existencia de dos ORFs, separadas en el genoma, que codifican proteínas de la cola, la más pequeña encargada de construir la principal proteína de la cola y la mayor de formar la proteína responsable de la adsorción al huésped (Lubbers y cols., 1995). Por el momento no se sabe si las dos proteínas de la cola de Cp-1 reflejan las dos diferentes funciones que desempeña la de ϕ 29, o si se trata de un caso similar al del fago c2.

Los conectores de los fagos son unas estructuras que unen la cabeza y la cola, y suelen estar formados por varias proteínas, la portal, las de terminación de la cabeza y las de inyección. La proteína portal es el sitio de entrada del DNA durante el proceso de empaquetamiento y el sitio de salida durante la inyección del DNA. Se ha descrito que los conectores de diferentes sistemas fágicos no relacionados presentan una morfología similar y comparten los mismos rasgos estructurales, lo cual indica que la conservación de esta estructura debería estar relacionada con la funcionalidad del conector y, por tanto, con la posibilidad de ensamblaje de las precabezas (Donate y Carrascosa, 1991). Aunque se ha visto que las proteínas de diferentes tipos de fagos que tienen funciones análogas en los viriones no presentan una homología de secuencia, excepto en fagos relacionados, una comparación de proteínas del conector de diferentes fagos muestra la existencia de una

región N-terminal cargada positivamente y de un pI básico (Rishovd y cols., 1994). En el caso de $\phi 29$, el conector está formado solamente por la proteína portal, que interviene en el empaquetamiento del DNA por medio de su unión al pRNA a través del sitio de unión a RNA que presenta su secuencia de aminoácidos (Donate y cols., 1993). Se ha demostrado que su dominio N-terminal cargado positivamente es el encargado de unirse al DNA para su posterior empaquetamiento en las precabezas a través de las cargas positivas que presenta y que interactúan con las negativas del complejo TP-DNA (Donate y cols., 1992), sin que exista una dependencia de alguna secuencia específica (Herranz y cols., 1986). El producto de la ORF10 de Cp-1 posee una gran similitud con la proteína del conector de $\phi 29$ a lo largo de toda la molécula, tiene un pI básico (5,28) y un N-terminal cargado positivamente, aspectos éstos característicos, como acabamos de ver, de las proteínas del conector de diferentes fagos. También se puede sugerir en su secuencia un dominio de unión a RNA (Kenan y cols., 1991), al igual que el existente en la de $\phi 29$, que se encargaría de unirse al pRNA de Cp-1, de manera análoga a lo propuesto para $\phi 29$.

Las proteínas de la encapsidación en los diferentes fagos son las encargadas de proporcionar la energía necesaria para la translocación del DNA al interior de las precabezas, por lo que todas ellas presentan motivos de unión a ATP. En el ejemplo de los DNAs lineales que se replican formando concatémeros, estas moléculas también digieren el DNA en sitios específicos, *cos* o *pac*, o en secuencias inespecíficas (Black, 1988). En el caso de $\phi 29$, cuyo DNA replica por "protein-priming", la proteína de la encapsidación hidroliza una molécula de ATP por cada dos bases empaquetadas (Guo y cols., 1987d). Esta proteína se une a las precabezas para realizar el empaquetamiento del DNA, a través de su unión al pRNA previamente unido al conector. Como se ha citado en el apartado 1.3 de Resultados, esta proteína presenta motivos de unión a ATP, Mg^{2+} (Guo y cols., 1987d) y RNA (Grimes y Anderson, 1990). De manera análoga, en la ORF20 de Cp-1 son claramente distinguibles los dos tipos de motivos de unión a ATP, A y B (Figura 8); además se puede sugerir la existencia de un motivo de unión a RNA y un probable sitio de unión a Mg^{2+} aunque, en este caso, los aminoácidos clave, cys e his, van consecutivos en lugar de estar separados por uno o dos residuos. Por otra parte, se ha visto durante el proceso de purificación de la proteína de encapsidación de $\phi 29$ (Guo y cols., 1986) que se trata de una proteína muy insoluble, lo que se ha relacionado con su alta hidrofobicidad: el contenido en F, Y, L, I, M, V y W es del 46% comparado con el 29% de media de las proteínas de *E.*

coli. Asimismo, en el caso de la de Cp-1, el contenido en estos aminoácidos alcanza un 40%.

El procesamiento proteolítico de las proteínas que participan en la morfogénesis viral es un fenómeno ampliamente generalizado y que tiene lugar en numerosos virus animales y bacterianos (Dougherty y Semler, 1993). Uno de los ejemplos más estudiados es el procesamiento de estas proteínas en el fago T4. En este caso, la proteasa gp21 codificada por el fago se ensambla en el interior de la precabeza y después, a causa de determinadas señales, se activa ella misma mediante proteólisis. La proteasa ahora activa, gp21*, digiere en pequeños fragmentos al menos diez de las proteínas estructurales que componen la precabeza para, finalmente, degradarse (Showe y cols., 1976a, b). La secuencia específica reconocida por esta proteinasa es Pho-Pho-Glu[↓], dónde Pho indica un aminoácido hidrofóbico, Glu, ácido glutámico y la flecha el sitio de corte (revisado por Black y Showe, 1983; Keller y cols., 1984). Otros fagos también presentan proteólisis de precursores de la cabeza, como P2 y lambda, las proteínas de la cola de lambda y T5, las del cuello y apéndices de ϕ 29, las de las fibras de la cola de T2 (revisado por Casjens y Hendrix, 1988) y, más recientemente, se ha descrito el procesamiento de la proteína de la cabeza de PR4 (Myung y cols., 1994), de HK97 (Duda y cols., 1995a, b) y en proteínas del fago A511 (Loessner y Scherer, 1995). No obstante, el origen y la especificidad de la correspondiente proteasa permanece todavía sin determinar en la mayoría de los casos (Casjens y Hendrix, 1988).

El ensamblaje de los bacteriófagos sigue generalmente un orden secuencial preciso (Casjens y Hendrix, 1988) de forma que, si alguno de los pasos no se produce o lo hace de manera inadecuada, el proceso no continúa sino que da lugar a la acumulación de un producto intermedio del proceso, que conduce a la formación de una partícula fágica aberrante. La proteína mayoritaria del virión de Cp-1 sufre el procesamiento de los 48 primeros aminoácidos de su N-terminal (Figura 17). Los resultados de nuestros experimentos demuestran que la proteína de la cabeza se procesa para formar el fago infectivo ya que, en el virión maduro, no se encuentra la forma de la proteína sin procesar y, en cambio, sí se encuentra en la banda superior que carece de DNA y que se aísla en la purificación del fago mediante centrifugación en CsCl. Como se ha comprobado mediante experimentos de "Western-blot" con anticuerpos frente a la proteína de la cabeza, en esta banda del gradiente aparecen las dos formas de la proteína de la cabeza (Figura 18A). La

explicación más plausible es que la proteína sin procesar no es capaz de formar una precabeza estable que empaquete el DNA y, simultáneamente, las precabezas correctamente formadas con la proteína procesada han sido incapaces de empaquetar y formar un virión maduro por algún defecto en las etapas posteriores. Esto estaría de acuerdo con los procesos de ensamblaje y empaquetamiento ampliamente estudiados en varios fagos (Casjens y Hendrix, 1988; Black, 1988). Nuestros resultados sugieren, además, que el producto de la ORF14 es responsable del procesamiento de la proteína de la cabeza de Cp-1, es decir, o bien codifica la proteasa encargada del proceso o es un factor que estimula una proteasa celular o la autoproteólisis de la proteína de la cabeza. Los resultados obtenidos sugieren que se trataría de una verdadera proteasa, ya que si fuera un factor que la induce o la activa, estaríamos en presencia de una enzima que reconoce una secuencia específica y que está presente en bacterias tan distantes como *E. coli* y neumococo, ya que el procesamiento se produce en las dos especies. Los resultados preliminares obtenidos en los clones de *E. coli* que expresan simultáneamente la ORF9, que codifica la proteína mayoritaria, y la ORF14 indican una baja eficiencia del procesamiento de la proteína. Ésto se podría deber a un nivel bajo de expresión de la proteasa en contraste con la hiperexpresión de la proteína de la cabeza o bien a la necesidad de algún otro factor, posiblemente también de origen fágico, que aumente la eficiencia del procesamiento, de forma similar a lo que se ha sugerido en el fago T4 (Hintermann y Kuhn, 1992), en los comovirus (Peters y cols., 1992a, b), o recientemente en adenovirus donde se requiere DNA y un péptido específico como cofactores de la reacción (Mangel y cols., 1993). El papel exacto de estos cofactores adicionales no se conoce, pero son imprescindibles en los ensayos de proteólisis. En cualquiera de los casos, parece que se trata de la primera proteasa codificada por un genoma fágico de pequeño tamaño, menos de 25 kb, ya que hasta ahora sólo se habían descrito proteasas codificadas por los genomas de los fagos lambda y T4, entre otros, todos ellos con genomas grandes. Recientemente, se ha postulado la existencia de una proteasa en el genoma, de aproximadamente 40 kb, del fago HK97 (Duda y cols., 1995a). Curiosamente, en el caso de ϕ 29 no se ha demostrado que codifique ninguna proteasa responsable del procesamiento de la proteína que forma los apéndices, ya que se ha sugerido que sería una proteína del huésped la que realiza la proteólisis (Tosi y cols., 1975). Por otra parte, en el fago PR4 (14,5 kb), relacionado con PRD1, la proteína mayoritaria de la cabeza es el precursor de otras dos proteínas de la cápsida, pero todavía

no se conoce nada acerca del origen de la proteasa encargada del procesamiento (Myung y cols., 1994).

Respecto a la especificidad del enlace hidrolizado, se han descrito una gran variedad de secuencias de reconocimiento entre las proteasas virales y del huésped, todas ellas implicadas en el procesamiento de distintas proteínas estructurales de los viriones (Casjens y Hendrix, 1988). Si se analiza la secuencia de corte de la proteasa de Cp-1 (Figura 17), se parece ligeramente a la que reconocen las serín-proteasas de flavivirus que tienen preferencia por cortar después de dos aminoácidos básicos y antes de un aminoácido de cadena corta como ala, ser o gly (Chambers y cols., 1990a, b). Por otra parte, la secuencia N-terminal de la proteína procesada presente en el virión (ala-thr-val) coincide con la correspondiente secuencia de reconocimiento de varias proteasas del tipo de las del fago T4 en el extremo N-terminal de la proteína liberada (Black y Showe, 1983), si bien la secuencia C-terminal del fragmento de 48 aminoácidos liberado en Cp-1 no revela ninguna similitud con la producida por las proteasas de T4.

En la búsqueda de motivos funcionales en la secuencia de aminoácidos del producto de la ORF14 con la base de datos PROSITE, se encontraron motivos característicos de las serín-proteasas que, aunque no están totalmente conservados, presentan los aminoácidos requeridos para su acción catalítica. Esto se correspondería con que la secuencia de corte se parece bastante a la reconocida por las serín-proteasas de flavivirus, como se ha citado anteriormente. Por lo tanto, muy probablemente, la ORF 14 codifica una serín-proteasa, aunque se necesitan experimentos adicionales para confirmar esta observación.

En la mayoría de los sistemas fágicos, existe un sistema dual de lisis en el que dos proteínas, holina y enzima lítica, son las encargadas de producir la lisis de la bacteria huésped para permitir la liberación de la progenie fágica. Los genes que codifican estas proteínas se encuentran normalmente agrupados en el genoma del fago, siendo el gen de la holina el que precede al de la enzima lítica, llegando a veces a solaparse el codón de terminación de una con el de iniciación del otro (Young, 1992; Steiner y cols., 1993; Birkeland, 1994; Ziermann y cols., 1994; Henrich y cols., 1995; Díaz y cols., 1996). No obstante, hay algunas excepciones en las que estos genes se encuentran alejados entre sí en el genoma como son los casos del fago T7, T3, bIL67, PRD1 o P1 (Studier, 1972; Miyazakiy cols., 1978; Schouler y cols., 1994; Bamford y cols., 1995; Schmidt y cols., 1996) e incluso hay casos en los que se han propuesto dos genes codificantes de presuntas

holinas, uno que precede al gen de la enzima lítica y otro alejado en el genoma, como en el fago c2 (Lubbers y cols., 1995) o el profago PBSX (Longchamp y cols., 1994). En los ejemplos estudiados hasta ahora, la transcripción de estos genes se produce en la fase tardía, con algunas excepciones como la lisozima de PRD1 que se transcribe en la etapa final de los genes tempranos (Bamford y cols., 1995), o la amidasa de T7 que es un gen de transcripción media que se expresa junto con los genes relacionados con el metabolismo del DNA, y que no parece ser necesario para la lisis del huésped (Studier, 1972).

Se ha demostrado que Cp-1 expresa una lisozima (García y cols., 1987), producto del gen 22 (*cpl1*), pero también codifica una segunda proteína implicada en el proceso de lisis, la holina, el producto del gen 21 (*cph1*). De forma similar a la mayoría de los sistemas descritos, el gen *cph1* precede al de la lisozima, estando solapados el codón de terminación de *cph1* con el de iniciación del gen *cpl1*. Estos genes se transcriben en la fase tardía a partir de dos promotores en tándem que preceden al gen *cph1*; además de este acoplamiento transcripcional, la secuencia del genoma predice la existencia de una horquilla de terminación de la transcripción en el extremo 3' del gen *cpl1* con un ΔG de 20,3 kcal/mol. Esto indicaría la existencia de un mecanismo de regulación transcripcional conjunta, que coordinaría, en el momento preciso, la acción de la holina y la subsiguiente liberación de la enzima lítica hacia su sustrato en la pared celular. En el caso de PRD1, en el que como se ha citado el gen de la holina no precede al de la lisozima, sintetizándose la holina en tiempos tardíos, esta disposición puede tener relación con una posible regulación de la lisis, porque aunque el gen de la lisozima se exprese pronto hasta que no se exprese el de la holina no va a poder producirse la lisis del huésped.

La regulación de los genes de holina puede también tener lugar a nivel postraducciona l pues, en algunos casos, se ha demostrado la existencia de un doble motivo de inicio de la traducción que se refleja en la acción antagonista de las dos proteínas sintetizadas siendo la de mayor tamaño inhibidora de la otra, que es la proteína letal. Por tanto, en los casos en que no existe este motivo debe existir otro sistema para regular la función de estas proteínas letales. En este sentido, recientemente se ha descrito en el fago P1 otra proteína distinta de la holina, la antiholina, que funciona como inhibidora de ésta (Schmidt y cols., 1996), y en el fago c2 se ha postulado la existencia de dos proteínas con las características estructurales de las holinas, una cuyo gen precede al de la enzima lítica mientras el otro gen está alejado de los anteriores en el genoma (Lubbers y cols., 1995) y la

expresión de ambas proteínas podría implicar, como en el caso de lambda, la existencia de un mecanismo de regulación entre ambas proteínas.

Los experimentos de expresión de los genes de la holina y de la lisozima de Cp-1, tanto en el sistema homólogo como en *E. coli*, han demostrado que ambas proteínas ejercen una función similar en las dos bacterias, si bien el comportamiento fisiológico y el resultado final de ambos sistemas fue bastante diferente. Un primer factor que explicaría los distintos efectos observados podría residir en la carencia de vectores con promotor regulable en neumococo, lo que obliga a clonar los genes líticos en unas condiciones especiales que limiten en lo posible su expresión constitutiva. En este sentido, se ha conseguido la viabilidad de clones de neumococo que expresan los genes de holina y lisozima de Cp-1, mediante la eliminación de su propio promotor y bajo el control del promotor del gen de la resistencia a Tc situado a 1,6 kb del extremo 5' de la holina. Con estas precauciones, el nivel de expresión de los genes líticos se reduce, aunque siguen notándose sus efectos pues la viabilidad de neumococo disminuye en media unidad logarítmica en cuatro horas, cuando sólo se expresa *cphI*, y cuando se expresan conjuntamente *cphI* y *cplI* desciende también media unidad logarítmica a las dos horas, antes de que el cultivo comience a lisarse. El efecto lítico sólo se observa, como era de esperar, cuando actúa la lisozima, aunque es interesante señalar que la lisis se produce también con la sola expresión del gen de lisozima sin intervención de la holina aunque, en este caso, el cultivo tarda seis horas en lisarse en vez de cuatro (Romero y cols., 1993). Esto apuntaría a que la lisozima de Cp-1 es capaz de alcanzar el sustrato, es decir de traslocarse, sin necesidad de holina, si bien la diferencia de efectividad entre ambas situaciones podría explicarse por la formación de un "agujero" de holina que aceleraría el paso de moléculas de lisozima con la consiguiente aceleración en la lisis del cultivo. La posibilidad de que la lisozima se transloque sin péptido señal no sería una propiedad peculiar de esta enzima pues ya se ha demostrado que la amidasa LytA es también capaz de alcanzar la pared sin un aparente péptido señal (Díaz y cols., 1989). De todas maneras, no es descartable que exista algún mecanismo particular de regulación en neumococo sin contrapartida en *E. coli*. En esta última especie, la acción de los genes líticos, una vez inducidos, es mucho más rápida y drástica que en neumococo, pues sólo con la expresión del gen de la holina se produce una pérdida de viabilidad de dos unidades logarítmicas en dos horas, mientras que la expresión de ambos genes produce un descenso de dos unidades logarítmicas después de 90 min antes de que el cultivo empiece a lisarse.

En este último caso la mortalidad del cultivo iba acompañada de lisis. Este mismo comportamiento, producto de la expresión de los dos genes líticos, se ha observado en *E. coli*, cuando los genes de los fagos $\phi 29$, $\phi LC3$, EJ-1 (Steiner y cols., 1993; Birkeland, 1994; Díaz y cols., 1996) se han clonado en este sistema. No obstante, en el caso del fago ϕadh , cuando se expresan sus genes de holina y enzima lítica en *E. coli*, sólo se produce detención del crecimiento del cultivo pero no lisis (Henrich y cols., 1995), lo que indica que no se produce una lesión importante en la membrana ya que si se añade un agente desestabilizante de la misma como el cloroformo, sí se produce la lisis del cultivo. Sin embargo, cuando se expresa la holina de ϕadh junto con la lisozima de $\phi 29$, sí se produce la lisis del cultivo. Esto sugiere que se requiere una señal adicional o algún otro mecanismo, ausente en *E. coli*, para que la holina de ϕadh posibilite el tránsito de la enzima lítica a través de la membrana y, por tanto, la regulación del proceso lítico mientras que en el sistema heterólogo estas interacciones regulatorias no deben ocurrir.

La elucidación de las características bioquímicas y fisiológicas de las holinas constituye uno de los grandes retos en la investigación actual sobre fagos. Prueba de ello es que la holina de lambda (producto del gen *S*) se viene estudiando hace muchos años y, hasta la fecha, ha sido difícil su purificación y no ha sido posible el conocimiento detallado de su anclaje topológico en la membrana. En este último caso, la visualización del “agujero” en la membrana interna por medio de microscopía electrónica no ha sido posible, probablemente debido a la alteración de las muestras durante su preparación para la observación microscópica. Por tanto, la metodología seguida en esta Memoria para demostrar la funcionalidad de la holina en el sistema homólogo ha sido, por un lado, la evidencia de la muerte de las células en cultivo y, por otro, la constatación de que las lesiones producidas en la membrana son las responsables de la liberación al periplasma de una enzima de localización exclusivamente citoplasmática como es la neumolisina (Johnson, 1977). Esta liberación es limitada en el caso del clon con *cphI*, algo superior en el clon que posee ambos genes, *cphI* y *cplI* y, en cualquier caso, claramente diferenciable del clon que sólo lleva el plásmido vector.

Como se ha descrito en la Introducción, la actividad de las enzimas líticas del sistema de neumococo requiere la presencia de colina en los ácidos teicoicos de la pared celular (Höltje y Tomasz, 1975a, b; Romero y cols., 1990a, b; García y cols., 1990; Díaz y cols., 1992a). Sin embargo, en esta Memoria se ha demostrado que CplI y LytA son

capaces de degradar paredes de bacterias Gram-negativas que carecen de colina (*E. coli*), como recientemente se había observado para Ejl, la enzima lítica del fago EJ-1 de neumococo (Díaz y cols., 1996). Estas cinéticas de degradación del sustrato observadas podían reflejar diferencias en la unión de las enzimas a paredes con estructuras y composiciones químicas diferentes que producen también diferencias en la eficiencia de degradación de la pared. Por lo tanto, se puede concluir que Ejl, LytA, Cpl1 y, probablemente, otras enzimas líticas de las paredes de neumococo, tienen una especificidad de sustrato más amplia de lo que se pensaba, siendo capaces de atacar no sólo paredes que contienen colina de bacterias Gram-positivas, *S. pneumoniae*, *C. acetobutylicum* (Croux y cols., 1993b) y *S. oralis* (Ronda y cols., 1988), sino también paredes de bacterias Gram-negativas que no contienen colina ni ácidos teicoicos. Esta versatilidad funcional podría explicarse en base a la poca complejidad estructural que presentan las paredes de las bacterias Gram-negativas ya que, como se ha demostrado *in vitro*, los fragmentos solubles de pared de neumococo carentes de colina también son buenos sustratos para la amidasa LytA (García-Bustos y Tomasz, 1987).

Las autolisinas bacterianas deben estar estrictamente reguladas durante la multiplicación celular. En el caso de la autolisina LytA, la regulación debe ser realizada post-traducionalmente porque el gen *lytA* se expresa constitutivamente y no se aprecian señales regulatorias en su unidad transcripcional (Díaz y García, 1990). La lisis de las células de *E. coli*, cuando expresan conjuntamente el gen *lytA* y el de la holina de Cp-1, indica que el que se dispare la acción de la amidasa se debe a una lesión de la membrana citoplasmática, lo que está de acuerdo con sugerencias previas que apuntaban que esta estructura jugaba un papel importante en el control de la actividad de la proteína LytA, tanto en *E. coli* como en neumococo (Díaz y cols., 1989, 1991). En concordancia con esto, se ha observado la liberación al medio de cultivo de lípidos de membrana y de macromoléculas que contienen colina antes de que se produzca la autólisis de los cultivos de neumococo en la fase estacionaria (Waks y Tomasz, 1978) y la necesidad de una membrana energizada para el control *in vivo* de la actividad de la autolisina en *B. subtilis* (Jolliffe y cols., 1981). Ronda y cols. (1977) propusieron la participación de un producto fágico en el disparo de la actividad de la autolisina de neumococo durante la liberación de la progenie fágica y, por lo tanto, una proteína tipo holina sensible al estado fisiológico de la célula que tendría una función análoga a los genes asesinos Gef a Relf (Poulsen y cols.,

1989) en *E. coli* podría ser la responsable del disparo de la actividad de la autolisina LytA. Un candidato hipotético para este gen podría ser el gen *cid* (Moreillon y cols., 1990).

Los genes de las enzimas líticas de los fagos Tuc2009, mv1, ϕ adh, entre otros, (Arent y cols., 1994; Boizet y cols., 1990; Henrich y cols., 1995) presentan una alta similitud con la de los fagos Cps. Por el contrario, los genes que codifican las holinas en estos sistemas no muestran similitudes significativas con la del fago Cp-1. Esta última observación se corresponde con los datos recogidos por Young y Bläsi (1995) en su reciente revisión. Estos datos sugieren que tales genes han evolucionado independientemente, como sucede en la evolución de estos genes en el fago lambda (Young, 1992) en donde cada región codificante define una unidad de intercambio. El análisis de la secuencias de los genes *cph1* y *cpl1* revela diferente uso de codones principalmente para gln, phe, ser y asn, lo que refuerza la idea de que los dos elementos del sistema de lisis deben haber evolucionado separadamente, como ya se ha sugerido para otros sistemas líticos de fagos como ϕ LC3 (Birkeland y cols., 1994) o EJ-1 (Díaz y cols., 1996). Por otro lado, a medida que se van secuenciando nuevos genomas fágicos van apareciendo genes que codifican presuntas holinas, que presentan grandes diferencias en sus secuencias pero mantienen la uniformidad en términos estructurales en lo que se refiere a la disposición de los residuos hidrofóbicos, hidrofílicos y cargados. Esto sugiere que los genes de las holinas de los diferentes fagos han evolucionado a partir de orígenes múltiples hacia una misma solución funcional para determinar el momento exacto en el que la enzima lítica de los fagos se tiene que liberar al periplasma para permitir la liberación de la progenie fágica. La extrema divergencia de secuencias entre las holinas debe reflejar requerimientos no estrictos a nivel de estructura primaria. No obstante, observaciones recientes muestran que la holina de EJ-1 sí presenta, a nivel de secuencia de aminoácidos, similitud (alrededor de un 50% de identidad) con las holinas propuestas para los fagos ϕ LC3, Tuc2009, ϕ 11 y el fago atemperado 392-A2 de *Lactococcus casei* (Arent y cols., 1994; Young y Bläsi, 1995; Díaz y cols., 1996; Herrero y cols., 1996), y también se encuentran similitudes a este nivel entre la holina de lambda y P22 (Young y Bläsi, 1995), lo que de algún modo introduce un elemento de duda sobre el origen evolutivo de las holinas y proporcionan una nueva luz de análisis funcional de estas proteínas basadas en la similitud de las estructuras primarias.

Por último, puesto que la función de la holina parece ser la de permitir el paso de la enzima lítica al periplasma por medio de una lesión inespecífica en la membrana plasmática,

permitiendo la liberación de componentes celulares, se puede sugerir que la holina puede tener una potencial aplicación en biotecnología para posibilitar la liberación de las proteínas no secretadas producidas por medio de la tecnología del DNA recombinante así como para su utilización en sistemas suicidas diseñados para la contención biológica.

3. MECANISMO DE INICIACIÓN DE LA REPLICACIÓN

El primer paso que tiene lugar durante la replicación del DNA tanto en replicones eucarióticos como procarióticos, es el desenrollamiento de la doble hélice del DNA en los orígenes de replicación para permitir la entrada de la maquinaria replicativa. Este hecho también se produce en los genomas lineales que tienen TP unida covalentemente a sus extremos y en este caso, ciertas proteínas de unión específica a DNA, como la proteína p6 en $\phi 29$ (Serrano y cols., 1990, 1993) o la proteína DBP de adenovirus (Stuvier y cols., 1992), parecen promover el desenrollamiento de los extremos del DNA. Después, los factores directamente implicados en la catálisis del paso de iniciación deben reconocer el origen monocatenario. De acuerdo con este modelo, hemos encontrado que extractos solubles parcialmente purificados de *S. pneumoniae* infectados con el bacteriófago Cp-1 pueden llevar a cabo la iniciación de la replicación de moldes de cadena simple, utilizando como cebador de la reacción, un residuo 3'-OH libre de una treonina de la TP. La iniciación de la replicación con TP sobre oligonucleótidos de cadena simple ya ha sido descrita en otros sistemas como $\phi 29$, PRD1 y en diferentes tipos de adenovirus (Méndez y cols., 1992; Caldentey y cols., 1993; Dobbs y cols., 1990; King y Van der Vliet, 1994).

El análisis mutacional de los extremos de Cp-1 usando oligonucleótidos sintéticos reveló que una de las tres timinas terminales (especialmente la segunda) disminuye fuertemente la eficiencia de iniciación. Por consiguiente, esas timinas deben de ser parte de una secuencia específica de DNA responsable del reconocimiento del molde en el origen de replicación abierto por el complejo TP-DNA polimerasa. El hecho de que las proteínas de Cp-1 sean activas sobre moléculas de DNA correspondientes a los extremos del DNA de $\phi 29$ se puede explicar por la gran similitud existente entre las secuencias de los extremos de ambos fagos (Figura 30A). Teniendo en cuenta que 6 de los 8 nt primeros son idénticos entre ambos sistemas, es lógico pensar que esta pequeña región de DNA esté implicada en la interacción con el complejo TP-DNA polimerasa. La distancia precisa de esta secuencia

específica al extremo 3' debe ser crítica para el reconocimiento del origen, tal y como sugiere el hecho de la extremadamente baja eficiencia de iniciación observada en las moléculas de DNA de Cp-1 que carecen de la T terminal. Estas conclusiones concuerdan esencialmente con las propuestas para la replicación en el caso de adenovirus tipo 2 (Dobbs y cols., 1990) y PRD1 (Yoo e Ito, 1991a, b) puesto que mutaciones en varias posiciones cercanas al extremo disminuyen o anulan totalmente la eficiencia de iniciación.

Los resultados obtenidos en nuestro trabajo indican que la iniciación de la replicación de Cp-1 está dirigida por el tercer nt del DNA y que se recuperan los dos primeros nucleótidos durante los primeros pasos de elongación. Aunque una secuencia específica de DNA puede jugar algún papel en el reconocimiento del origen, la selección del sitio de iniciación parece estar impuesta por el complejo TP-DNA polimerasa, independientemente de la secuencia implicada. Esta hipótesis se apoya en el hecho de que las proteínas de Cp-1 también inician en tercera posición utilizando como moldes moléculas de DNA correspondientes a los orígenes de $\phi 29$ o a los orígenes de Cp-1 que carecen de una de las tres timinas terminales. La iniciación interna en genomas con TP fue descrita por primera vez en la replicación del fago $\phi 29$ (Méndez y cols., 1992). El DNA de $\phi 29$ contiene una reiteración de tres timinas en sus dos extremos 3', y la formación del complejo TP-dAMP es dirigida por el segundo nucleótido. Sin embargo, el primer nucleótido se recupera durante la elongación, pero se necesita una reiteración de al menos dos timinas en los extremos 3' para que la replicación sea eficiente. Todos estos datos sugirieron que el complejo de iniciación TP-dAMP se deslizaría hacia atrás para situar el residuo dAMP frente a la primera T del molde. El hecho de que todos los genomas con TP tengan algún tipo de reiteración en los extremos de sus DNAs llevó a proponer que la iniciación interna y el consiguiente deslizamiento hacia atrás o "sliding-back" es un mecanismo común a todos los genomas que usan el sistema de "protein-priming" para la iniciación de la replicación (Méndez y cols., 1992).

En los últimos años, esta predicción se ha demostrado en varios sistemas. El análisis mutacional del origen del fago PRD1 (cuya reiteración consiste en cuatro cicteínas) demostró que la sustitución de la primera, segunda, o tercera C no elimina la formación del complejo TP-dGMP (Yoo e Ito, 1991b). Caldentey y cols. (1993) demostraron que la iniciación es dirigida por la cuarta C, y se propuso la recuperación de la secuencia de los extremos del DNA por medio de pasos sucesivos de translocación hacia atrás. No obstante,

no se consiguieron pruebas directas que demostraran la existencia de estos sucesivos pasos, ya que la DNA polimerasa no elonga *in vitro* los oligonucleótidos de cadena simple. Por otra parte, en muchos serotipos de adenovirus, la reiteración terminal consiste en la repetición de un trinucleótido (3'-GTAGTA). El análisis mutacional del adenovirus tipo 4 reveló que la sustitución simultánea de la cuarta y la séptima G del molde impide completamente la formación del complejo TP-dCTP (Harris y Hay, 1988). Un análisis similar en adenovirus tipo 2 mostró que la sustitución de la primera G no afecta a la formación de dicho complejo, mientras que una sustitución adicional en la cuarta G, lo inhibe (Dobbs y cols., 1990). Además, recientemente, King y Van der Vliet (1994) han demostrado que el intermedio de replicación de adenovirus TP-CAT se sintetiza usando como molde el trinucleótido GTA en las posiciones 4-6, y el complejo TP-CAT "salta" hacia las posiciones 1-3 del molde para comenzar la elongación. Esta variación del mecanismo de "sliding-back" llamada "jumping-back" permite la regeneración *in vitro* de delecciones de uno o dos nucleótidos de los extremos del DNA.

Después de la iniciación en el tercer nucleótido, hay dos posibilidades para explicar la recuperación de los extremos del DNA de Cp-1: o dos pasos sucesivos de "sliding-back" o un simple salto de la tercera a la primera posición. En ambos casos, el complejo DNA polimerasa-TP que sirve como cebador debe reconocer el origen, situando la tercera T del molde frente al sitio de unión a dNTPs de la DNA polimerasa. Entonces, el dATP sería seleccionado como sustrato para catalizar la formación de TP-dAMP. En el modelo de "sliding-back" paso a paso (Figura 34A), la entrada de una segunda molécula de dAMP debería estar acoplada a la traslocación de TP-dAMP una posición más atrás en el molde, situándose el dAMP inicial frente a segunda T del molde. De la misma manera, la incorporación de la tercera molécula de dAMP, ocurriría acoplada a la traslocación del producto TP-(dAMP)₂, completamente emparejada con el extremo del DNA. Después de estos pasos especiales de traslocación, tendría lugar el proceso normal de elongación. Este doble paso de "sliding-back" podría explicar que los moldes con mutaciones en la primera o tercera Timinas no fueran completamente replicados (Figura 34C). Aunque el oligonucleótido ori(29) T1C puede ser un eficiente molde para la iniciación, solo un paso de "sliding-back" puede ocurrir sin que se produzcan desapareamientos. Entonces, o bien el complejo TP-AA se disocia de la cadena molde en el segundo "sliding-back" o bien la elongación se produce sin la recuperación del nucleótido terminal. En el caso del

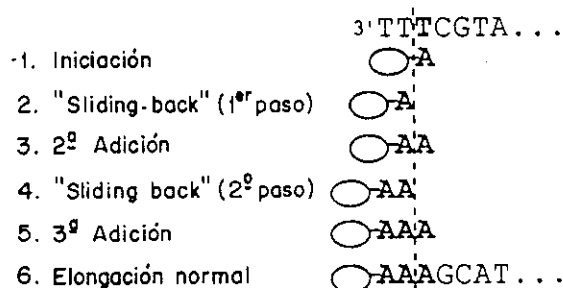
oligonucleótido ori(29) T3C, incluso la primera traslocación sería incorrecta en el apareamiento de bases y las moléculas eventualmente replicadas carecerían de los dos nucleótidos terminales. En ambos casos, estas moléculas de DNA con reiteraciones terminales incompletas no se utilizarán eficientemente en el nuevo ciclo de replicación (como se deduce de la extremada baja eficiencia de iniciación detectada con el oligonucleótido $\Delta T1$, que carece de la T terminal).

Por otra parte, la posibilidad de un único salto hacia atrás desde la tercera a la primera T del molde (Figura 34B) no puede ser excluida, y de hecho podría explicar el curioso resultado en el que los productos de replicación del oligonucleótido ori(29) T2C fueron similares en longitud a los obtenidos con la secuencia salvaje (Figura 34C). En particular, la mutación T2C en el extremo del DNA de Cp-1 produce la repetición de un dinucleótido (3'- TCTC...) que podría facilitar un paso de preelongación y el consecuente "jumping-back" del intermedio TP-AG, de manera similar al mecanismo de adenovirus (Figura 34C).

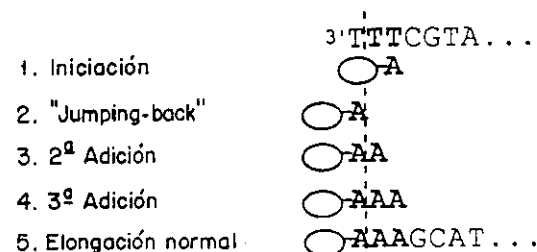
Los datos presentados indican que el sistema es lo suficientemente flexible para permitir tanto el "sliding-back" paso a paso como el "jumping-back", y parece difícil predecir qué mecanismo prefiere el origen salvaje. Sin embargo, dos diferentes argumentos nos conducen a favor del modelo del "sliding-back" paso a paso. Primero, la iniciación en la tercera posición del DNA podría ser una consecuencia de una incapacidad estructural de la DNA polimerasa de Cp-1 para usar como molde el primer y segundo nucleótido terminal. La clave del hecho del modelo de "sliding-back" paso a paso es que sólo la tercera T es requerida para dirigir la incorporación de los tres residuos dAMP consecutivos, mientras que en el modelo de "jumping-back" se requeriría un uso eficiente de ambos, del segundo y del tercer nucleótido como molde. La segunda razón proviene del hecho de que el modelo de "sliding-back" paso a paso podría contribuir a la fidelidad de la replicación de los extremos del DNA de Cp-1, puesto que los nucleótidos unidos a la TP en la reacción de iniciación serían comprobados por apareamiento de bases durante los dos pasos de traslocación. Por el contrario, el mecanismo de "jumping-back", que implica traslocación directa del tercer al primer nucleótido, podría aumentar la fijación de errores en el segundo nucleótido.

La iniciación de la replicación del DNA de los fagos con TP puede considerarse un paso de alto riesgo pues, al menos en el caso de $\phi 29$ la DNA polimerasa comete muchos más

A "Sliding-back" paso a paso



B "Jumping-back"



C Orígenes mutados

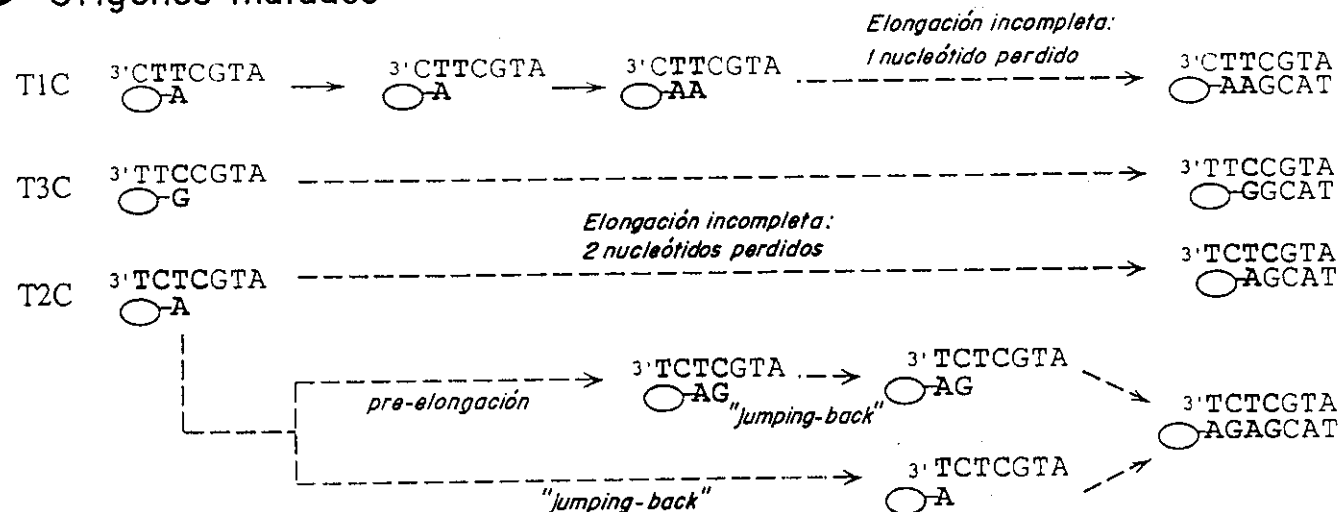


Figura 34. Modelos del mecanismo para la transición entre la iniciación y la elongación de la replicación. A) Modelo de "sliding-back" paso a paso. La integridad de los extremos del DNA se consigue mediante dos pasos sucesivos de "sliding-back". El tercer nucleótido del extremo 3' del DNA se usó como molde durante todo el proceso (tres veces en total). B) Modelo de "jumping-back". La integridad de los extremos del DNA se consigue mediante un salto desde la tercera posición a la primera. C) Efecto de las mutaciones en las reiteraciones terminales. Se indican las diferentes posibilidades para explicar los productos de elongación parcial obtenidos con los oligonucleótidos indicados en la figura. Las flechas continuas indican las reacciones que ocurren como se indica en el panel A de la figura. Las flechas discontinuas indican las vías alternativas para la replicación de los orígenes mutados.

errores en la inserción de dNMPs a la TP que en un paso normal de elongación, es decir, cataliza la inserción TP-dAMP con baja fidelidad y además, la actividad exonucleasa 3'→5' de la DNA polimerasa no puede actuar sobre el primer nucleótido unido a la TP (Esteban y cols., 1993). Por esta razón cabría esperar una alta frecuencia de mutación en el extremo del DNA. Sin embargo, tanto ϕ 29 como los fagos relacionados mantienen la misma secuencia invariable: 3'-TTTCAT, por lo que, en este caso el "sliding-back" provee un paso no exonucleolítico para la fidelidad de la iniciación (Méndez y cols., 1992; Blasco y cols., 1992; Salas y cols., 1992). En resumen, la replicación del DNA de Cp-1 presenta las ventajas de un mecanismo sofisticado de transición entre la iniciación por "protein-priming" y una elongación normal que contribuye de manera no exonucleolítica a prevenir las mutaciones en alguno de los tres nucleótidos terminales situados en el origen de replicación.

V. CONCLUSIONES

De los resultados presentados en esta Memoria se pueden deducir las siguientes conclusiones:

- 1.- El genoma del bacteriófago Cp-1 de *S. pneumoniae* contiene 19.345 pb y el análisis de su secuencia revela la existencia de 29 marcos abiertos de lectura, 23 en la cadena superior y las 6 restantes en la inferior.
- 2.- El análisis comparativo de las secuencias de aminoácidos deducidas de las ORFs muestra que 10 proteínas del fago Cp-1 son similares a las del fago ϕ 29 con una identidad de alrededor del 28%, que alcanza aproximadamente un 80% de similitud cuando también se consideran los aminoácidos conservativos.
- 3.- Se ha elaborado el mapa de transcripción del fago Cp-1, comprobándose que los RNAs transcritos tempranamente están localizados en los extremos del genoma, mientras los tardíos se sitúan en la parte central. Entre los genes caracterizados como tempranos se han localizado los que codifican la TP y la DNA polimerasa, en tanto que entre los tardíos se encuentran los que codifican las proteínas morfogenéticas y las proteínas implicadas en la lisis.
- 4.- Se han caracterizado en el genoma de Cp-1 quince regiones promotoras, cuatro de ellas en tándem. Los genes tempranos inician la transcripción a partir de tres sitios diferentes mientras las otras siete regiones promotoras son responsables de la transcripción de los genes tardíos.
- 5.- Se ha demostrado la existencia de una región promotora situada en el extremo derecho del genoma, orientada en dirección hacia el extremo del genoma que sintetiza un RNA pequeño, implicado, probablemente, en el empaquetamiento del DNA.
- 6.- Se han clonado y expresado en *E. coli* trece ORFs del genoma de Cp-1, entre las que se encuentran las que codifican la DNA polimerasa, la TP, la proteína mayoritaria de la cabeza, el conector, la cola, la holina y la lisozima.

7.- La proteína mayoritaria de la cabeza se procesa por su extremo N-terminal, para formar la cápsida del virión.

8.- La proteasa responsable del procesamiento de la proteína mayoritaria de la cabeza de Cp-1 está codificada por el fago. Nuestros datos sugieren que la ORF14 sintetiza esta proteasa.

9.- La expresión de la holina de Cp-1, tanto en *E. coli* como en neumococo, provoca una alta tasa de mortalidad en los cultivos, pero no va acompañada de lisis celular.

10.- La expresión conjunta de los genes que codifican la holina y la lisozima de Cp-1, en *E. coli* y en neumococo, conduce a la lisis celular.

11.- La iniciación de la replicación del DNA de Cp-1 está dirigida por el tercer nucleótido del genoma y las dos primeras bases se recuperan durante el proceso de replicación, mediante dos pasos consecutivos de deslizamiento hacia atrás o mecanismo de "sliding-back".

12.- La especificidad del nucleótido que inicia la replicación viene dada por las características propias de la DNA polimerasa y la TP de Cp-1 y no por la secuencia del DNA que se replica.

VI. BIBLIOGRAFÍA

- Ackermann, H.W., Elzanowski, A., Fobo, G. y Stewart, G. 1995. Relationships of tailed phages: a survey of protein sequence identity. *Arch. Virol.* **140**: 1.871-1.884.
- Altman, E., Altman, R., Garret, R., Grimaila, R. y Young, R. 1983. *S* gene product: identification and membrane localization of a lysis control protein. *J. Bacteriol.* **155**: 1.130-1.137.
- Anderson, D. y Bodley, J.W. 1990. Role of RNA in bacteriophage ϕ 29 DNA packaging. *J. Struct. Biol.* **104**: 70-74.
- Anderson, D. y Reilly, B. 1993. Morphogenesis of bacteriophage ϕ 29. *En: Bacillus subtilis and other Gram-positive bacteria: biochemistry, physiology, and molecular genetics.* ASM publications. pp. 859-867.
- Appleyard, R.K. 1954. Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. *Genetics* **39**: 440.
- Arendt, E. K., Dali, C., Fitzgerald, G.F. y Van de Guchte, M. 1994. Molecular characterization of lactococcal bacteriophage Tuc2009 and identification and analysis of the genes encoding lysin, a putative holin, and two structural proteins. *Appl. Envir. Microbiol.* **60**: 1.875-1.883.
- Austrian, R. 1981. Pneumococcus: the first one hundred years. *Rev. Infect. Dis.* **3**: 183-189.
- Austrian, R. 1986. Some aspects of the pneumococcal carrier state. *J. Antimicrob. Chemoter.* **18**: 35-45.
- Avery, O.T., MacLeod, C.M. y McCarthy, M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. I. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **79**: 137-157.
- Bacot, C. M., y Reeves, R. H. 1991. Novel tRNA gene organization in the 16S-23S intergenic spacer of the *Streptococcus pneumoniae* rRNA gene cluster. *J. Bacteriol.* **173**: 4.234-4.236.
- Bailey, S., Wichitwechkarn, J., Johnson, D., Reilly, B.E., Anderson, D. y Bodley J.W. 1990. Phylogenetic analysis and secondary structure of the *Bacillus subtilis* bacteriophage RNA required for DNA packaging. *J. Biol. Chem.* **265**: 22.365-22.370.
- Bamford, D.H., Caldentey, J. y Bamford, J.K.H. 1995. Bacteriophage PRD1: a broad host range dsDNA tectivirus with an internal membrane. *Adv. Virus Res.* vol. **45**: 281-319.
- Bamford, J. K. H., Hänninen, A-L., Pakula, T.M., Ojala, P. M., Kalkkinen, N., Frilander, M. y Bamford, D. H. 1991. Genome organization of membrane-containing bacteriophage PRD1. *Virology* **183**: 658-676.
- Bamford, J. K. H. y Mindich, L. 1984. Characterization of the DNA-protein complex at the termini of the bacteriophage PRD1 genome. *J. Virol.* **50**: 309-315.
- Barany, F. y Tomasz, A. 1980. Genetic transformation of *Streptococcus pneumoniae* by heterologous plasmid deoxyribonucleic acid. *J. Bacteriol.* **144**: 698-709.
- Bernard, A., Zaballo, A., Salas, M. y Blanco, L. 1987. Structural and functional relationships between prokaryotic and eukaryotic DNA polymerase. *EMBO J.* **6**: 4.219-4.225.
- Bernheimer, H.P. 1977. Lysogeny in pneumococci freshly isolated from man. *Science* **195**: 66-68.
- Bernheimer, H.P. 1979. Lysogenic pneumococci and their bacteriophages. *J. Virol.* **138**: 618-624.
- Berry, A.M., Lock, R.A., Hansman, D. y Paton, J.C. 1989. Contribution of autolysin to virulence of *Streptococcus pneumoniae*. *Infect. Immun.* **57**: 2.324-2.330.
- Bienkowska-Szewczyk, K. y Taylor, A. 1980. Murein transglycosylase from phage λ lysate purification and properties. *Biochem. Biophys. Acta* **615**: 489-496.
- Birkeland, N-K. 1994. Cloning, molecular characterization, and expression of the genes encoding the lytic functions of lactococcal bacteriophage ϕ LC3: a dual lysis system of modular design. *Can. J. Microbiol.* **40**: 658-665.
- Birnboim, H. C. y Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1.513-1.523.
- Bjornsti, M.A., Reilly, B. y Anderson, D. 1983. Morphogenesis of bacteriophage *Bacillus subtilis*: oriented and quantized in vitro packaging of DNA protein gp3. *J. Virol.* **45**: 383-396.
- Blackburn, E.H. Telomerases. *Annu. Rev. Biochem.* **61**: 113-129.
- Black, L. 1988. DNA packaging in dsDNA bacteriophages. *En: The bacteriophages.* R. Calendar (ed). 2, pp. 321-373. Plenum, New York.

- Black, L. y Showe, M. 1983. Morphogenesis of the T4 head. Bacteriophage T4, pp. 219-245. ASM publications. Washintong.
- Blasco, M.A., Esteban, J.A., Méndez, J., Blanco, L. y Salas, M. 1992. Structural and functional studies on ϕ 29 DNA polymerase. *Chromosoma* 102: 32-38.
- Bläsi, U., Nam, K., Hartz, D., Gold, L. y Young, R. 1989. Dual translational initiation sites control function of the lambda S gene. *EMBO J.* 8: 3.501-3.510.
- Bläsi, U., Halfman, G. y Lubitz W. 1984. *En: Microbial. Cell Wall Synthesis and autolysis* (Nombela, C., de) pp. 213-218. Elsevier Science Publishers, Amsterdam.
- Bläsi, U., Chang, C.Y., Zagotta, M., Nam, K. y Young, R. 1990. The lethal λ S gene encodes its own inhibitor. *EMBO J.* 9: 981-989
- Boizet, B., Lahbib-Masais, Y., Dupont, L., Ritzenthaler, P. y Mata, M. 1990. Cloning, expressiion and sequence analysis of an endolysin-encoding gene of *Lactobacillus bulgaricus* bacteriophage mv1. *Gene* 94: 61-67.
- Bolívar, F. 1978. Construction and characterization of new cloning vehicles. III. Derivatives of plasmid pBR322 carrying unique *EcoRI* sites for selection of *EcoRI* generated recombinant DNA molecules. *Gene* 4: 121-136.
- Bolívar, F., Rodríguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. y Falcow, S. 1977. Construction and characterization of new cloning vehicles. *Gene* 2: 95-113.
- Bonovich, M.T. y Young, R. 1991. Dual start motif in two lambdoid S genes unrelated to lambda S gene. *J. Bacteriol.* 173: 2.897-2.905.
- Botstein, D., Chan, R.K. y Waddel, C.H. 1972. Genetics of bacteriophage P22 II. Gene order and gene function. *Virology* 49: 268-282.
- Boyer, H. y Roulland-Dussoix, D. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* 41: 459-274.
- Braithwite, D.K. e Ito, J. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* 21: 787-802.
- Bradford, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254.
- Burns, C.M., Chan, H.L.B. y DuBow, M. 1990. *In vitro* maturation and encapsidation of the DNA of transposable Mu-like phage 108. *Proc. Natl. Acad. Sci. USA* 87: 6.092-6.096.
- Caldentey, J., Blanco, L., Bamford, D.H. y Salas, M. 1993. *In vitro* replication of bacteriophage PRD1 DNA. Characterization of the protein-primed initiation site. *Nucleic Acids Res.* 21: 3.725-3.730.
- Caldentey, J., Hanninen, A.L. y Bamford, D.H. 1994. Gene XV of bacteriophage PRD1 encodes a lytic enzyme with muramidase activity. *Eur. J. Biochem.* 225: 341-346.
- Camacho, A., Jiménez, F., de la Torre, J., Carrascosa, J.L., Mellado, R.P., Vázquez, C., Viñuela, E. y Salas, M. 1977. Assembly of *Bacillus subtilis* phage ϕ 29. Mutants of the cistrons coding for the structural proteins. *Eur. J. Biochem.* 73: 39-55.
- Campbell, A. 1988. Phage evolution and speciation. *En: The Bacteriophages. R. Calendar (ed).* 1, pp. 1-14. Plenum, New York.
- Carafa, Y., Brody, E. y Thermes, C. 1990. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.* 216: 835-858.
- Casjens, S. y Hendrix, R.W. 1988. Control mechanisms in dsDNA bacteriophage assembly. *En: The Bacteriophages. R. Calendar (ed).* 1, pp. 15-92.
- Chambers, T.J., McCourt, D.W. y Rice, C.M. 1990a. Production of yellow fever virus in infected cells: identification of discrete polyprotein species and analysis of cleavage kinetics using region specific polyclonal antisera. *Virology* 177: 159-174.
- Chambers, T.J., Hahn, C.S., Galler, R. y Rice, C.M. 1990b. Flavivirus genome organization, expression, and replication. *Annu. Rev. Microbiol.* 44: 649-688.

- Chandry, P.S., Davidson, B.E. y Hillier, A.J. 1994. Temporal transcription map of the *Lactococcus lactis* bacteriophage sk1. *Microbiology* **140**: 2.251-2.261.
- Chang, P-C y Cohen, S. 1994. Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. *Science* **265**: 952-954.
- Cheetham, B.F. y Katz, M. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol. Microbiol.* **18**: 201-208.
- Chou, P. y Fasman, G.D. 1978. Empirical prediction of protein conformation. *Annu. Rev. Biochem.* **47**: 251-276.
- Croux, C., Ronda, C., López, R. y García, J.L. 1993a. Interchange of functional domains switches enzyme specificity: construction of a chimeric pneumococcal-clostridial cell wall lytic enzyme. *Mol. Microbiol.* **9**: 1.019-1.025.
- Croux, C., Ronda, C., López, R. y García, J.L. 1993b. Role of the C-terminal domain of the lysozyme of *Clostridium acetobutylicum* ATCC 824 in a chimeric pneumococcal-clostridial cell wall lytic enzyme. *FEBS Lett.* **336**: 111-114.
- Daubert, S.D. y Bruening, G. 1984. Detection of genome-linked proteins of plants and animal viruses. *Methods Virol.* **8**: 347-379.
- DeMartini, M., Halegoua, S. y Inouye, M. 1975. Lysozymes from bacteriophages T3 and T5. *J. Virol.* **16**: 459-461.
- Devereux, J., Haeberli, P. y Smithies, O. 1984. A comprehensive set of sequence analysis program for the VAX. *Nucleic Acids Res.* **12**: 387-395.
- D'Hèrelle, F. 1917. Sur un microbe invisible antagonists des *Bacillus dysenteriques*. *Compt. Rend.* **165**: 373.
- Díaz, E. y García, J.L. 1990. Characterization of the transcription unit encoding the major pneumococcal autolysin. *Gene* **90**: 157-162.
- Díaz, E., García, E., Ascaso, C., Méndez, E., López, R. y García, J.L. 1989. Subcellular localization of the major pneumococcal autolysin: a peculiar mechanism of secretion in *Escherichia coli*. *J. Biol. Chem.* **264**: 1.238-1.244.
- Díaz, E., López, R. y García, J.L. 1990. Chimeric phage-bacterial enzymes: a clue to the molecular evolution of genes. *Proc. Natl. Acad. Sci. USA* **87**: 8.125-8.129.
- Díaz, E., López, R. y García, J.L. 1991. Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. *J. Biol. Chem.* **266**: 5.464-5.471.
- Díaz, E., López, R. y García, J.L. 1992a. EJ-1, a temperate bacteriophage of *Streptococcus pneumoniae* with a *Myoviridae* morphotype. *J. Bacteriol.* **174**: 5.516-5.525.
- Díaz, E., Munthali, M., Lünsdorf, H., Hölftje, J.V. y Timmis, K. 1996. The two-step lysis system of pneumococcal bacteriophage EJ-1 is functional in Gram-negative bacteria: triggering of the major pneumococcal autolysin in *Escherichia coli*. *Mol. Microbiol.* **19**: 667-681.
- Díaz, A., Pons, M.E., Lacks, S.A. y López, P. 1992b. *Streptococcus pneumoniae* DNA polymerase I lacks 3'-to-5' exonuclease activity: Localization of the 5'-to -3' exonucleolytic domain. *J. Bacteriol.* **174**: 2.014-2.024.
- Dobbs, L., Zhao, L.J., Sripad, G. y Padmanabhan, R. 1990. Mutational analysis of single-stranded DNA templates active in the *in vitro* initiation assay for adenovirus DNA replication. *Virology* **178**: 43-51.
- Donate, L., y Carrascosa, J.L. 1991. Characterization of a versatile *in vitro* DNA-packaging system based on hybrid $\lambda/\phi 29$ proheads. *Virology* **182**: 534-544.
- Donate, L., Valpuesta, J.M., Mier, C., Rojo, F. y Carrascosa, J.L. 1993. Characterization of an RNA-binding domain in the bacteriophage $\phi 29$ connector. *J. Biol. Chem.* **268**: 20.198-20.204.
- Donate, L., Valpuesta, J.M., Rocher, A., Méndez, E., Rojo, F., Salas, M y Carrascosa, J.L. 1992. Role of the amino-terminal domain of bacteriophage $\phi 29$ connector in DNA binding and packaging. *J. Biol. Chem.* **267**: 10.919-10.924.
- Dougherty, W.G. y Semler, B.L. 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol. Rev.* **57**: 781-822.

- Dower, W.J., Miller, J.F y Ragsdale, C.W. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* 16: 6.127-6.145.
- Dubos, R. J. 1937. Mechanism of the lysis of pneumococci by freezing and thawing, bile, and other agents. *J. Exp. Med.* 66: 101-112.
- Duda, R.L., Martinic, K. y Hendrix, R.W. 1995a. Genetic basis of bacteriophage HK97 prohead assembly. *J. Mol. Biol.* 247: 636-647.
- Duda, R.L., Martinic, K., Xie, Z. y Hendrix, R.W. 1995b. Bacteriophage HK97 head assembly. *FEMS Microbiol. Rev.* 17: 41-46.
- Dupont, L., Vasla, A., Mikkonen, M., Boizet-Bonhoure, B., Alatossava, T. y Ritzenthaler, P. 1993. CenBank accession number Z26590.
- Eppler, K., Wyckoff, E., Goates, J., Parr, R. y Casjens, S. 1991. Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. *Virology* 183: 519-538.
- Escarmís, C., García, P., Méndez, E., López, R., Salas, M. y García, E. 1985. Inverted terminal repeats and terminal proteins of the genome of pneumococcal phages. *Gene* 36: 341-348.
- Escarmís, C., Gómez, A., García, E., Ronda, C., López, R. y Salas, M. 1984. Nucleotide sequence at the termini of the DNA of *Streptococcus pneumoniae* phage Cp-1. *Virology* 133: 166-171.
- Escarmís, C. y Salas, M. 1981. Nucleotide sequence at the termini of the DNA of the *Bacillus subtilis* phage ϕ 29. *Proc. Natl. Acad. Sci. USA* 78: 1.446-1.450.
- Esteban, J.A., Salas, M. y Blanco, L. 1993. Fidelity of ϕ 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.* 268: 2.719-2.726.
- Fenoll, A., Martínez-Suarez, J., Muñoz, R., Casal, J. y García, J.L. 1990. Identification of atypical strains of *Streptococcus pneumoniae* by a specific DNA probe. *Eur. J. Clin. Microbiol. Infect. Dis.* 9: 396-401.
- Ferreras, M., Gavilanes, J.G. y García-Segura, J.M. 1993. A permanent Zn^{2+} reverse staining method for the detection and quantification of proteins in polyacrylamide gels. *Anal Biochem.* 213: 206-212.
- Freimuth, P.I. y Ginsberg, H.S. 1986. Codon insertion mutants of the adenovirus terminal protein. *Proc. Natl. Acad. Sci. USA* 83: 7.816-7.820.
- García, J.L., García, E., Arrarás, A., García, P., Ronda, C. y López, R. 1987. Cloning, purification, and biochemical characterization of the pneumococcal bacteriophage Cp-1 lysin. *J. Virol.* 61: 2.573-2.580.
- García, E., García, J.L., García, P., Arrarás, A., Sánchez-Puelles, J.M. y López, R. 1988. Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Proc. Natl. Acad. Sci. USA* 85: 914-918.
- García, P., García, J.L., García, E. y López, R. 1986a. Nucleotide sequence and expression of the pneumococcal autolysin gene from its own promoter in *Escherichia coli*. *Gene* 43: 265-272.
- García, P., García, J.L., García, E. y López, R. 1989. Purification y characterization of the autolytic glycosidase of *Streptococcus pneumoniae*. *Biochem. Biophys. Res. Commun.* 158: 251-256.
- García, P., García, J.L., García, E., Sánchez-Puelles, J.M. y López, R. 1990. Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene* 86: 81-88.
- García, E., García, J. L., Ronda, C., García, P. y López, R. 1985a. Cloning and expression of the pneumococcal autolysin gene in *Escherichia coli*. *Mol. Gen. Genet.* 201: 225-230.
- García, P., García, E., Ronda, C., López, R. y Tomasz, A. 1983a. A phage associated murein-hydrolase in *Streptococcus pneumoniae* infected with bacteriophage Dp-1. *J. Gen. Microbiol.* 129: 489-497.
- García, P., García, E., Ronda, C., Tomasz, A. y López, R. 1983b. Inhibition of lysis by antibody against phage-associated lysin and requirement of choline-residues in the cell wall for progeny phase release in *Streptococcus pneumoniae*. *Curr. Microbiol.* 8: 137-140.

- García, E., Gómez, A., Ronda, C., Escarmís, C. y López, R. 1983c. Pneumococcal bacteriophage Cp-1 contains a protein bound to the 5' termini of its DNA. *Virology* 128: 92-104.
- García, P., Hermoso, J. M., García, J.A., García, E., López, E. y Salas, M.. 1986b. Formation of a covalent complex between the terminal protein of pneumococcal bacteriophage Cp-1 and 5'-dAMP. *J. Virol.* 58: 31-35.
- García, P., López, R., Ronda, C., García, E. y Tomasz, A. 1983d. Mechanism of phage-induced lysis in pneumococci. *J. Gen. Microbiol.* 129: 479-487.
- García, P., Méndez, E., García, E., Ronda, C. y López, R. 1984. Biochemical characterization of a murein hydrolase induced by bacteriophage Dp-1 in *Streptococcus pneumoniae*: comparative study between bacteriophage-associated lysin and the host amidase. *J. Bacteriol.* 159: 793-796.
- García, E., Ronda, C., García, J.L. y López, R. 1985b. A rapid procedure to detect the autolysin phenotype in *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* 29: 77-81.
- García, E., Ronda, C., García, P. y López, R. 1986c. Studies on the replication of bacteriophage Cp-1 DNA in *Streptococcus pneumoniae*. *Microbiología SEM* 2: 115-120.
- García-Bustos, J.F. y Tomasz, A. 1987. Teichoic acid-containing muropeptides from *Streptococcus pneumoniae* as substrates for the pneumococcal autolysin. *J. Bacteriol.* 169: 447-453.
- Garrett, J., Bruno, C., y Young, R. 1990. Lysis protein S of phage lambda functions in *Saccharomyces cerevisiae*. *J. Bacteriol.* 172: 7.275-7.277.
- Garrett, J. y Young, R. 1982. Lethal action of bacteriophage λ S gene. *J. Virol.* 44: 886-892.
- Geis, A. 1992. Cloning and DNA sequence analysis of a lysin gene of lactococcal bacteriophage P001. *En: Annual Report 1991. Federal Dairy Research Centre, Kiel, Germany*, pp. B94.
- Gibson, T. J. 1984. Tesis Doctoral. Universidad de Cambridge, Inglaterra.
- Griffith, F. 1928. The significance of pneumococcal types. *J. Hyg.* 27: 113-159.
- Grimes, S. y Anderson, D. 1989. Cleaving the prohead RNA of bacteriophage ϕ 29 alters the *in vitro* packaging of restriction fragments of DNA-gp3. *J. Mol. Biol.* 209: 101-108.
- Grimes, S. y Anderson, D. 1990. RNA dependence of the bacteriophage ϕ 29 DNA packaging ATPase. *J. Mol. Biol.* 215: 559-566.
- Guo, P., Bailey, S., Bodley, J.W. y Anderson, D. 1987a. Characterization of the small RNA of the bacteriophage ϕ 29 DNA packaging machine. *Nucleic Acids Res.* 15: 7.081-7.090.
- Guo, P., Erickson, S. y Anderson, D. 1987b. A small viral RNA is required for *in vitro* packaging of bacteriophage ϕ 29 DNA. *Science* 236: 690-694.
- Guo, P., Grimes, S. y Anderson, D. 1986. A defined system for *in vitro* packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage ϕ 29. *Proc. Natl. Acad. Sci. USA* 83: 3.505-3.509.
- Guo, P., Peterson, C. y Anderson, D. 1987c. Initiation events in *in vitro* packaging of bacteriophage ϕ 29 DNA-gp3. *J. Mol. Biol.* 197: 219-228.
- Guo, P., Peterson, C. y Anderson, D. 1987d. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage ϕ 29. *J. Mol. Biol.* 197: 229-236.
- Hager, P.W. y Rabinowitz, J.C. 1985. Translational specificity in *Bacillus subtilis*. *En: The molecular biology of the bacilli*, vol. II: pp. 1-32.
- Hanahan, D. 1983. Studies on transformation of *E. coli* with plasmids. *J. Mol. Biol.* 166: 557-580.
- Harley, C.B. y Reynolds, R.P. 1987. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.* 15: 2.343-2.361.
- Harris, M.P.G. y Hay, R.T. 1988. DNA sequences required for the initiation of adenovirus type 4 DNA replication *in vitro*. *J. Mol. Biol.* 201: 57-67.
- Hatfield, L. y Hearing, P. 1993. The NFIII/OCT-1 binding site stimulates adenovirus DNA replication *in vivo* and is functionally redundant with adjacent sequences. *J. Virol.* 67: 3.931-3.939.
- Hatfull, G.F. y Sarkis, G.J. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol. Microbiol.* 7: 395-405.

- Hayashida, M., Watanabe, K., Muramatsu, T. y Aki Goto, M. 1987. Further characterization of PL-1 phage associated *N*-acetylmuramidase of *Lactobacillus casei*. J. Gen. Microbiol. 133: 1.343-1.349.
- Henrich, B., Binishofer, B. y Bläsi, U. 1995. Primary structure and functional analysis of the lysis genes of *Lactobacillus gasseri* bacteriophage ϕ adh. J. Bacteriol. 177: 723-732.
- Hermoso, J.M. y Salas, M. 1985. Protein p3 is linked to the DNA of phage ϕ 29 through a phosphoester bond between serine and 5'-dAMP. Proc. Natl. Acad. Sci. USA 77: 6.425-6.428.
- Hermoso, J. M., Méndez, E., Soriano, F. y Salas, M. 1985. Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage ϕ 29. Nucleic Acids Res. 13: 7.715-7.728.
- Herranz, L., Salas, M. y Carrascosa, J.L. 1986. Interaction of the bacteriophage ϕ 29 connector protein with the viral DNA. Virology 155: 289-292.
- Herrero, M. 1996. Caracterización del bacteriófago temperado 393-A2 de *Lactobacillus casei*. Memoria de Tesis Doctoral. Universidad de Oviedo.
- Hinnebusch, J. y Tilly, K. 1993. Linear plasmids and chromosomes in bacteria. Mol. Microbiol. 10: 917-922.
- Hintermann, E. y Kuhn, A. 1992. Bacteriophage T4 gene encodes two proteins essential for phage maturation. Virology 189: 474-482.
- Hirel, H., Schmitter, J.M., Dessen, P., Fayat, G. y Blanquet, S. 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA 86: 8.247-8.251.
- Hirochika, H., Nakamura, K. y Sakaguchi, K. 1984. A linear DNA plasmid from *Streptomyces rochei* with an inverted terminal repetition of 614 base pairs. EMBO J. 3: 761-766.
- Hirochika, H. y Sakaguchi, K. 1982. Analysis of linear plasmids isolated from *Streptomyces*: association of protein with the ends of the plasmid DNA. Plasmid 7: 59-65.
- Höltje, J.V. y Tomasz, A. 1975a. Lipoteichoic acid: a specific inhibitor of autolysin activity in pneumococcus. Proc. Nat. Acad. Sci. USA 72: 1.690-1.694.
- Höltje, J.V. y Tomasz, A. 1975b. Specific recognition of choline residues in the cell wall teichoic acid by the *N*-Acetylmuramyl-L-alanine amidase of pneumococcus. J. Biol. Chem. 250: 6.072-6.076.
- Höltje, J.V. y Tomasz, A. 1976. Purification of the pneumococcal *N*-acetylmuramyl-L-alanine amidase to biochemical homogeneity. J. Biol. Chem. 251: 4.199-4.207.
- Howard, L.V. y Gooder, H. 1974. Specificity of the autolysin of *Streptococcus (Diplococcus) pneumoniae*. J Bacteriol. 117: 796-804.
- Inouye, M., Arnheim, N. y Sternglanz, R. 1973. Bacteriophage T7 lysozyme is an *N*-acetylmuramyl-L-alanine amidase. J.Biol. Chem. 248: 7.247-7.252.
- Inouye, M. y Delihias, N. 1988. Small RNAs in the prokaryotes: A growing list of diverse roles. Cell 53: 5-7.
- Inouye, S. e Inouye, M. 1985. Up-promoter mutations in the *lpp* gene of *Escherichia coli*. Nucleic Acids Res. 13: 3.101-3.110.
- Inouye, M. y Tsugita, T. 1968. Amino acid sequence of T2 phage lysozyme. J. Mol. Biol. 37: 213-219.
- Ishino, Y., Iwasaki, H., Kato, I. y Shinagawa, H. 1994. Amino acid sequence motifs essential to 3'→5' exonuclease activity of *Escherichia coli* DNA polymerase II. J. Biol. Chem. 269: 14.655-14.660.
- Jiménez, F., Camacho, A., de la Torre, J., Viñuela, E. y Salas, M. 1977. Assembly of *Bacillus subtilis* phage ϕ 29. 2. Mutants in the cistrons coding for the non-structural proteins. Eur. J. Biochem. 73: 57-72.
- Johnson, M.K. 1977. Cellular location of pneumolysin. FEMS Microbiol. Lett. 2: 243-245.
- Jolliffe, L.K., Doyle, R.J. y Streips, U.N. 1981. The energized membrane and cellular autolysis in *Bacillus subtilis*. Cell 25: 753-763.

- Josslin, R. 1970. The lysis mechanism of phage T4: mutants affecting lysis. *Virology* **40**: 719-726.
- Josslin, R. 1971. Physiological studies of the *t* gene defect in T4-infected *Escherichia coli*. *Virology* **44**: 101-107.
- Joyce, C.M. 1991. Can DNA polymerase I serve as a model for other polymerases?. *Curr. Opin. Struc. Biol.* **1**: 123-129.
- Karawya, E., Swack, J. y Wilson, S.H. 1983. Improved conditions for activity gel analysis of DNA polymerase catalytic polypeptides. *Anal. Biochem.* **135**: 318-325.
- Karnik, S. y Billeter, M. 1983. Lysis function of RNA bacteriophage Q β is mediated by the maturation (A₂) protein. *EMBO J.* **2**: 1.521-1.526.
- Kawamura, T. y Shockman, G.D. 1983. Evidence for the presence of a second peptidoglycan hydrolase in *Streptococcus faecium*. *FEMS Microbiol. Lett.* **19**: 65-69.
- Keen, C.L., Mendelovitz, S., Cohen, G. y Aharonowitz, Y. y Roy, K.L. 1988. Isolation and characterization of a linear DNA plasmid from *Streptomyces clavuligerus*. *Mol. Gen. Genet.* **212**: 172-176.
- Keller, B., Senstag, C., Kellenberger, E. y Bickel, T. 1984. Gene 68, a new bacteriophage T4 gene which codes for the 17 Kd prohead core protein is involved in head size determination. *J. Mol. Biol.* **179**: 415-
- Kemble, R.J. y Thompson, R.D. 1982. S1 and S2, the linear mitochondrial DNAs present in a male sterile line of maize, posses terminally attached proteins. *Nucleic Acids Res.* **10**: 8.181-8.190.
- Kenan, D.J., Query, C.C. y Keene, J.D. 1991. RNA recognition: towards identifying determinants of specificity. *TIBS* **16**: 214-220.
- Kikuchi, Y., Hirai, K. y Hishinuma, F. 1984. The yeast linear DNA killer plasmids, pGKL1 and pGKL2, possess terminally attached proteins. *Nucleic Acids Res.* **12**: 5.685-5.692.
- Kilpper-Bälz, R., Wenzig, P. y Schleifer, K.H. 1985. Molecular relationships and classification of some viridans streptococci as *S. oralis* and emended description of *Streptococcus oralis* (Bridge and Sneath 1982). *Int. J. Syst. Bacteriol.* **35**: 482-488.
- Kim, Y.T., Tabor, S., Bortner, C., Griffith, J.D. y Richardson, C.C. 1992. Purification and characterization of the bacteriophage T7 gene 2.5 protein. A singled stranded DNA-binding protein. *J. Biol. Chem.* **267**: 15.022-15.031.
- King, T.C., Sirdeskumkh, R. y Schlessinger, D. 1986. Nucleolytic processing of ribonucleic acid transcripts in procaryotes. *Microbiol. Rev.* **50**: 428-451.
- King, A.J. y Van der Vliet, P.C. 1994. A precursor terminal protein -trinucleotide intermediate during initiation of the adenovirus DNA replication: regeneration of molecular ends *in vitro* by a jumping back mechanism. *EMBO J.* **13**: 5.786-5.792.
- Kornberg, A. 1980. *DNA replication*. San Francisco: Freeman.
- Kornberg, A. 1982. *DNA replication. Supplement*. San Francisco: Freeman.
- Kornberg, A. y Baker, A. 1992. *DNA replication*. San Francisco: Freeman.
- Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A. y Hayward, R.S. 1993. The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J. Mol. Biol.* **232**: 406-418.
- Kushner, S.R. 1978. *En: Genetic engineering* (Boyer, H.B. y Nicosia, S., eds.) pp. 17. Elsevier/North Holland, Amsterdam.
- Kyogoku, K. y Sekiguchi, J. 1996. Cloning and sequencing of a new holin-encoding gene of *Bacillus licheniformis*. *Gene* **168**: 61-65.
- Lacks, S.A. y Neuberger, M. 1975. Membrane location of a deoxyribonuclease implicated in the genetic transformation of *Diplococcus pneumoniae*. *J. Bacteriol.* **124**: 1.321-1.329.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophag T4. *Nature* **227**: 680-685.
- Lee, C.S. y Guo, P. 1994. A highly sensitive system for the *in vitro* assembly of bacteriophage ϕ 29 of *Bacillus subtilis*. *Virology* **202**: 1.039-1.042.

- Leung, Y.C. y Errington, J. 1995. Characterization of an insertion in the phage $\phi 105$ genome that blocks host *Bacillus subtilis* lysis and provides strong expression of heterologous genes. *Gene* **154**: 1-6.
- Lin, Y-S., Kieser, H.M., Hopwood, D.A. y Chen, C. 1993. The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol. Microbiol.* **10**: 923-933.
- Loessner, M.J. y Scherer, S. 1995. Organization and transcriptional analysis of the *Listeria* phage A511 late gene region comprising the major capsid and tail sheath protein genes *cps* and *tsh*. *J. Bacteriol.* **177**: 6.601-6.609.
- Loessner, M.J., Wendlinger, G. y Scherer, S. 1995. Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes. *Mol. Microbiol.* **16**: 1.231-1.241.
- Longchamp, P.F., Mauel, C. y Karamata, D. 1994. Lytic enzymes associated with defective prophages of *Bacillus subtilis*: sequencing and characterization of the region comprising the N-acetylmuramoyl-L-alanine amidase gene of prophage PBSX. *Microbiology* **140**: 1.855-1.867.
- López, R., García, E., García, P. y García, J.L. 1995. Architecture and domain interchange of the pneumococcal cell wall lytic enzymes. *En: Genetics of Streptococci, Enterococci and Lactococci*. vol **85**, pp. 273-281.
- López, R., García, J. L., García, E., Ronda, C. y García, P. 1992. Structural analysis and biological significance of the cell wall lytic enzymes of *Streptococcus pneumoniae* and its bacteriophage. *FEMS Microbiol. Lett.* **100**: 439-448.
- López, R., García, E. y Ronda, C. 1981. Bacteriophages of *Streptococcus pneumoniae*. *Rev. Infect. Dis.* **3**: 212-223.
- López, R., García, P., Ronda, C. y García, E. 1982. Genetics and biology of pneumococcal phages. *Microbiology* pp. 141-144.
- López, R., García, J.L., Ronda, C., García, E. y López, R. 1994. Enzimas líticas del sistema de neumococo. *Microbiol. SEM* **10**: 13-18.
- López, P., Martínez, S., Díaz, A., Espinosa, M. y Lacks, S.A. 1989. Characterization of the *polA* gene of *Streptococcus pneumoniae* and comparison of the DNA polymerase I it encodes to homologous enzymes from *Escherichia coli* and phage T7. *J. Biol. Chem.* **264**: 4.255-4.263.
- López, R., Ronda, C., García, P., Escarmís, C. y García, E. 1984. Restriction cleavage maps of the DNAs of *Streptococcus pneumoniae* bacteriophages containing protein covalently bound to their 5' ends. *Mol. Gen. Genet.* **197**: 67-74.
- López, R., Ronda, C., Tomasz, A., Portolés, A. 1977. Properties of "diplophage": a lipid containing bacteriophage. *J. Virol.* **24**: 201-210.
- Lubbers, M.W., Waterfield, N.R., Beresford, T.P.J., Le Page, R.W.F. y Jarvis, A. W. 1995. Sequencing and analysis of the prolate-headed lactococcal bacteriophage $\phi 29$ genome and identification of the structural genes. *Appl. Envir. Microbiol.* **61**: 4.348-4.356.
- Mangel, W.F., McGrath, W.J., Toledo, D.L. y Anderson, C.W. 1993. Viral DNA and a viral peptide are cofactors of adenovirus virion proteinase activity. *Nature* **361**: 274-275.
- Martin, B. y Claverys, J.P. 1991. *En: Genetics and molecular biology of Streptococci, Lactococci and Enterococci*. ASM.
- McDonnell, M., Ronda, C. y Tomasz, A. 1975. "Diplophage": a bacteriophage of *Diplococcus pneumoniae*. *Virology* **63**: 577-582.
- Méndez, J. 1995. Replicación del DNA del bacteriófago $\phi 29$. Caracterización del mecanismo de iniciación con proteína terminal. Memoria de Tesis Doctoral. Universidad Autónoma de Madrid.
- Méndez, J., Blanco, L., Esteban, J.A., Bernard, A. y Salas, M. 1992. Initiation of $\phi 29$ DNA replication occurs at the second 3' nucleotide of the linear template: a sliding-back mechanism for protein primed DNA replication. *Proc. Natl. Acad. Sci. USA* **89**: 9.579-9.583.
- Mermod, N., Ramos, J.L., Lehrbach, P.R. y Timmis, K.N. 1986. Vector for regulated expression of cloned genes in a wide range of Gram-negative bacteria. *J. Bacteriol.* **167**: 447-454.
- Mindich, L. y Lehman, J. 1979. Cell wall lysis as a component of the bacteriophage $\phi 6$ virion. *J. Virol.* **30**: 489-496.

- Miyazaki, J.I., Ryo, Y., Fujisawa, H. y Minagawa, T. 1978. Mutation in bacteriophage T3 affecting host cell lysis. *Virology* **89**: 327-329.
- Montang, D., Riede, I., Eschbach, M.L., Degen, M. y Henning, U. 1987. Receptor-recognizing proteins of T-even type bacteriophages. Constant and hypervariable regions and unusual case of evolution. *J. Mol. Biol.* **196**: 165-174.
- Moreillon, P., Markiewicz, Z., Nachman, S. y Tomasz, A. 1990. Two bactericidal targets for penicillin in pneumococci: autolysis-dependent and autolysis-independent killing mechanisms. *Antimicrob. Agents Chemother.* **34**: 33-39.
- Morgenroth, J. y Kaufmann, M. 1912. Arzneifestigkeit bei bakterien (pneumokokken). *Z. Immuniatatsforsch. Exp. Ther.* **15**: 610-624.
- Morrison, D. A., y B. Jaurin. 1990. *Streptococcus pneumoniae* possesses canonical *Escherichia coli* (sigma 70) promoters. *Mol. Microbiol.* **4**: 1.143-1.152.
- Mullan W.M.A. y Crawford, R.J.M. 1985. Partial purification and some properties of ϕ C2(W) lysin, a lytic enzyme produced by phage-infected cells of *Streptococcus lactis* C2. *J. Dairy Res.* **52**: 123-138.
- Murray, C.L. y Rabinowitz, J.C. 1982. Nucleotide sequences of transcription and translation initiation regions in *Bacillus* phage ϕ 29 early genes. *J. Biol. Chem.* **257**: 1.053-1.062.
- Myung, H., Vanden Boom, T. y Cronan, J.E. 1994. The major capsid protein of the lipid containing bacteriophage PR4 is the precursor of two other capsid proteins. *Virology* **198**: 17-24.
- Nam, K., Bläsi, U., Zagotta, M.T. y Young, R. 1989. Conservation of the dual-start motif in P22 lysis gene regulation. *J. Bacteriol.* **72**: 204-211.
- Neufeld, F. 1900. Ueber eine spezifische bakterolytische wirkung der galle. *Z. Hyg. Infekt.* **34**: 454-464.
- Neufeld, F. 1902. Ueber die agglutination der pneumokokken und über die theorieen der agglutination. *Z. Hyg. Infekt.* **40**: 54-72.
- Oram, J.D. y Reiter, B. 1965. Phage-associated lysins affecting group N and group D streptococci. *J. Gen. Microbiol.* **40**: 57-70.
- Paces, V., Vlcek, C., Smarda, J., Zadrazil, S. y Fucik, V. 1987. Tolerated variations in a genome: of closely related *Bacillus* phages PZA, ϕ 29 and ϕ 15- a review. *Gene* **54**: 155-165.
- Paces, V., Vlcek, C. y Urbanek, P. 1986. Nucleotide sequence of the late region of *Bacillus subtilis* phage PZA, a close relative of ϕ 29. *Gene* **44**: 107-114.
- Paillard, M., Sederoff, R. R. y Levings, C. S. 1985. Nucleotide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize. *EMBO J.* **4**: 1.125-1.128.
- Pakula, T.M., Savilahti, H. y Bamford, D.H. 1989. Comparison of the amino acids sequence of the lytic enzyme from broad-host-range bacteriophage PRD1 with sequences of other cell-wall-peptidoglycan lytic enzymes. *Euro. J. Biochem.* **180**: 149-152.
- Parsons, B.L. y Pickup, D.J. 1990. Transcription of orthopoxvirus telomeres at late times during infection. *Virology* **175**: 69-80.
- Payne, J. y Gasson, M. 1990. Cloning and characterization of a lysin gene from a *Lysteria* bacteriophage. *En: NATO ASI Series. H* **65**: pp. 427-431.
- Peters, S.A., Voorhorst, G.B., Wellink, J. y van Kammen, A. 1992a. Processing of Vpg-containing polyproteins encoded by the B-RNA from cowpea mosaic virus. *Virology* **191**: 90-97.
- Peters, S.A., Voorhorst, G.B., Wery, J., Wellink, J. y van Kammen, A. 1992b. A regularoty role for the 32K protein in proteolytic processing of cowpea mosaic virus polyproteins. *Virology* **191**: 81-90.
- Pettit, S.C., Horwitz, M.S. y Engler, J.A. 1989. Mutations of the precursor to the terminal protein of the adenovirus serotypes 2 and 5. *J. Virol.* **63**: 5.244-5.250.
- Platteeuw, C. y De Vos, W.M. 1992. Location, characterization and expression of lytic enzyme-encoding gene, *lytA*, of *Lactococcus lactis* bacteriophage ϕ US3. *Gene* **118**: 115-120.
- Poulsen, L.K., Lasen, N.W., Molin, S. y Andersson, P. 1989. A family of genes encoding a cell-killing function may be conserved in all Gram-negative bacteria. *Mol. Microbiol.* **3**: 1.463-1.472.

- Rane, L. y Subbarow, Y. 1940. Nutritional requirements of the pneumococcus 1. Growth factors for types I, II, V, VII, VIII. *J. Bacteriol.* **40**: 695-704.
- Reader, R.W. y Siminovitch, L. 1971. Lysis defective mutants of bacteriophage lambda: genetics and physiology of S cistron mutants. *Virology* **43**: 607-622.
- Reid, R.J.D., Bodley, J.W. y Anderson, D. 1994a. Identification of bacteriophage ϕ 29 prohead RNA domains necessary for *in vitro* DNA-gp3 packaging. *J. Biol. Chem.* **269**: 9.084-9.089.
- Reid, R.J.D., Zhang, F., Benson, S. y Anderson, D. 1994b. Probing the structure of bacteriophage ϕ 29 prohead RNA with specific mutations. *J. Biol. Chem.* **269**: 18.656-18.661.
- Rekosh, D.M.K., Russell, W.C. y Bellett, A.J.D. 1977. Identification of a protein linked to the ends of adenovirus DNA. *Cell* **11**: 283-295.
- Rennell, D. y Poteete, A.R. 1985. Phage P22 lysis genes: nucleotide sequences and functional relationships with T4 and lambda genes. *Virology* **143**: 280-289.
- Rietsch, A. y Bläsi, U. 1993. Non-specific hole formation in the *Escherichia coli* inner membrane by λ S proteins is independent of cellular *secY* and *secA* functions and of the proportion of membrane acidic phospholipids. *FEMS Microbiol. Lett.* **107**: 101-106.
- Rishvov, S., Marvik, O.J., Jacobsen, E. y Lindqvist, B. 1994. Bacteriophage P2 and P4 morphogenesis: identification and characterization of the portal protein. *Virology* **200**: 744-751.
- Rogers, H.J., Perkins, H.R. y Ward, J.B. 1980. Microbial cell walls and membranes. Chapman and Hall, London.
- Romero, A., López, R. y García, P. 1990a. Characterization of the pneumococcal bacteriophage HB-3 amidase: cloning and expression in *Escherichia coli*. *J. Virol.* **64**: 137-142.
- Romero, A., López, R. y García, P. 1990b. Sequence of the *Streptococcus pneumoniae* bacteriophage HB-3 amidase reveals high homology with the major host autolysin. *J. Bacteriol.* **172**: 5.064-5.070.
- Romero, A., López, R., Lurz, R. y García, P. 1990c. Temperate bacteriophages of *Streptococcus pneumoniae* that contain protein covalently bound to the 5' ends of their DNAs. *J. Virol.* **64**: 5.149-5.155.
- Romero, A., López, R. y García, P. 1993. Lytic action of cloned pneumococcal phage lysis genes in *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* **108**: 87-92.
- Ronda, C., García, J.L. y López, R. 1988. Characterization of genetic transformation in *Streptococcus oralis* NCTC 11427. Expression of pneumococcal amidase in *S. oralis* using a new shuttle vector. *Mol. Gen. Genet.* **215**: 53-57.
- Ronda, C., García, J. L. y López, R. 1989. Infection of *Streptococcus oralis* NCTC 11427 by pneumococcal phages. *FEMS Microbiol. Lett.* **65**: 187-192.
- Ronda, C., López, R. y García, E. 1981. Isolation and characterization of a new bacteriophage, Cp-1, infecting *Streptococcus pneumoniae*. *J. Virol.* **40**: 551-559.
- Ronda, C., López, R., Gómez, A. y García, E. 1983. Protease-sensitive transfection of *Streptococcus pneumoniae* with bacteriophage Cp-1 DNA. *J. Virol.* **48**: 721-730.
- Ronda, C., López, R., Tapia, A. y Tomasz, A. 1977. Role of the pneumococcal autolysin (murein hydrolase) in the release of the progeny bacteriophage and in the bacteriophage-induced lysis of the host cell. *J. Virol.* **21**: 366-374.
- Ronda, C., García, J.L., García, E., Sánchez-Puelles, J.M. y López, R. 1987. Biological role of the pneumococcal amidase. *Eur. J. Biochem.* **164**: 621-624.
- Roof, W.D. y Young, R. 1995. ϕ X174 lysis requires *slyD*, a host gene which is related to the FKBP family of peptidyl-prolyl *cis-trans* isomerases. *FEMS Microbiol. Rev.* **17**: 213-218.
- Sabelnikov, A.G., Greenberg, B. y Lacks, S.A. 1995. An extended -10 promoter alone directs transcription of the *DpnII* operon of *Streptococcus pneumoniae*. *J. Mol. Biol.* **250**: 144-155.
- Sable, S. y Lortal, S. 1995. The lysins of bacteriophages infecting lactic acid bacteria. *Appl. Microbiol. Biotechnol.* **43**: 1-6.
- Saedi, M., Garvey, K.J. e Ito, J. 1987. Cloning and purification of a unique lysozyme produced by *Bacillus* phage ϕ 29. *Proc. Natl. Acad. Sci. USA* **84**: 955-958.
- Sahn, F. 1990. Pneumococcus and influenza. *Lancet* **335**: 898-901.

- Salas, M. 1988. Phages with protein attached to the DNA ends. *En: The Bacteriophages*. R. Calendar (ed). 1, pp. 169-191. Plenum, New York.
- Salas, M. 1991. Protein-priming of DNA replication. *Annu. Rev. Biochem.* 60: 39-71.
- Salas, M., Méndez, J., Esteban, J.A., Bernard, A., Soengas, M.S., Lázaro, J.M., Blasco, M.A. y Blanco, L. 1992. Protein-primed replication of bacteriophage DNA. *En: Mossbach Colloquium: DNA replication and the cell cycle*. Springer-Verlag, vol. 43: 27-34.
- Salas, M. y Rojo, F. 1993. Replication and transcription of bacteriophage ϕ 29 DNA. *En: Bacillus subtilis and other Gram-positive bacteria: biochemistry, physiology, and molecular genetics*, ASM publications, pp. 843-857.
- Sambrook, J., Fritsch, E. F. y Maniatis, T. 1989. Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor. New York.
- Sancar, A., Hack, A.M. y Rupp, W.D. 1979. Simple method for identification of plasmid-coded proteins. *J. Bacteriol.* 137: 692-693.
- Sánchez-Puelles, J.M., Ronda, C., García, J.L., García, P., López, R. y García, E. 1986a. Searching for autolysin functions. Characterization of a pneumococcal mutant deleted in the *LyfA* gene. *Eur. J. Biochem.* 158: 289-293.
- Sánchez-Puelles, J.M., Ronda, C., García, E., Méndez, E., García, J.L. y López, R. 1986b. A new peptidoglycan hydrolase in *Streptococcus pneumoniae*. *FEMS Microbiol. Lett.* 35: 163-166.
- Sandmeier, H. 1994. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibers. *Mol Microbiol.* 12: 343-350.
- Sanger, F., Nicklen, S. y Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74: 5.463-5.467.
- Sanz, J.M. y García, J.L. 1990. Structural studies of the lysozyme coded by the pneumococcal phage Cp-1. Conformational changes induced by choline. *Eur. J. Biochem.* 187: 409-416.
- Schmidt, C., Velleman, M. y Arber, W. 1996. Three functions of bacteriophage P1 involved in cell lysis. *J. Bacteriol.* 178: 1.099-1.104.
- Schön, P., Schrot, G., Wanner, G., Lubitz, W. y Witte, A. 1995. Two-stage model for integration of the lysis protein E of ϕ X174 into the cell envelope of *Escherichia coli*. *FEMS Microbiol. Rev.* 17: 207-212.
- Schouler, C., Ehrlich, S.D. y Chopin, M.C. 1994. Sequence and organization of the lactococcal prolate-headed bIL67 phage genome. *Microbiology* 140: 3.061-3.069.
- Serrano, M., Gutiérrez, C., Salas, M. y Hermoso, J.M. 1993. Superhelical path of the DNA in the nucleoprotein complex that activates the initiation of phage ϕ 29 DNA replication. *J. Mol. Biol.* 230: 248-259.
- Serrano, M., Salas, M. y Hermoso, J.M. 1990. A novel nucleoprotein complex at a replication origin. *Science* 248: 1.012-1.016.
- Seto H. y Tomasz, A. 1975. Protoplast formation and leakage of intramembrane cell components: Induction by competence activator substance of pneumococci. *J. Bacteriol.* 121: 344-353.
- Shearman, C., Underwood, H., Jury, K. y Gasson, M. 1989. Cloning and DNA sequence analysis of a *Lactococcus* bacteriophage lysin gene. *Mol. Gen. Genet.* 218: 214-221.
- Shearman, C.A., Jury, K.L. y Gasson, M.J. 1994. Controlled expression and structural organization of a *Lactococcus lactis* bacteriophage lysin encoded by two overlapping genes. *Appl. Envir. Microbiol.* 60: 3.063-3.073.
- Sheehan, M.M., García, J.L., López, R. y García, P. 1996. Analysis of the catalytic domain of the lysin of the lactococcal bacteriophage Tuc2009 by chimeric gene assembling. *En prensa*.
- Shilavy, T.J., Berman, M.L. y Enquist, L.W. 1984. *En: Experiments in genes fusions*. Cold Spring Harbor laboratory Press. Cold Spring Harbor. New York.
- Shiue, S.-Y., Hsieh, J.-C. e Ito, I. 1991. Mapping of the DNA linking tyrosine residue of the PRD1 terminal protein. *Nucleic Acids Res.* 19: 3.805-3.810.
- Showe, M., Isobe, E. y Onorato, L. 1976a. Bacteriophage T4 prehead proteinase. I. Purification and properties of a bacteriophage enzyme which cleaves the capsid precursor proteins. *J. Mol. Biol.* 107: 35-

- Showe, M., Isobe, E. y Onorato, L. 1976b. Bacteriophage T4 prehead proteinase. II. Its cleavage from the product of the 2I and regulation in phage infected cells. *J. Mol. Biol.* **107**: 55-
- Smart, J.E. y Stillman, B.W. 1982. Adenovirus terminal protein precursor. Partial amino acid sequence and the site of covalent linkage to virus DNA. *J. Biol.Chem.* **257**: 13.499-13.506.
- Soengas, M.S., Esteban, J.A., Lázaro, J.M., Bernard, A., Blasco, M.A., Salas, M. y Blanco, L. 1992. Site-directed mutagenesis at the EXO III motif of ϕ 29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. *EMBO J.* **11**: 4.227-4.237.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments by gel electrophoresis. *J. Mol. Biol.* **98**: 507-517.
- Spiecher, D.W. 1994. Methods and strategies for the sequence analysis of proteins on PVDF membranes. *Methods* **6**: 262-273.
- Stahl, F. y Murray, N. 1966. The evolution of gene clusters and genetic circularity in microorganisms. *Genetics* **58**: 569.
- Steiner, M. y Bläsi, U. 1993. Charged amino-terminal amino acids affect the lethal capacity of lambda lysis protein S07 and S105. *Mol. Microbiol.* **8**: 525-533.
- Steiner, M., Lubitz, W. y Bläsi, U. 1993. The missing link in phage lysis of Gram positive bacteria: gene 14 of *Bacillus subtilis* phage ϕ 29 encodes the functional homolog of lambda S protein. *J. Bacteriol.* **175**: 1.038-1.042.
- Stillman, B.W. 1983. The replication of adenovirus DNA with purified proteins. *Cell* **35**: 7-9.
- Studier, F. W. 1972. Bacteriophage T7. Genetic and biochemical analysis of this simple phage gives information about basic genetic processes. *Science* **176**: 367-376.
- Studier, F.W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219**: 37-44.
- Studier, F.W. y Moffat, B.A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. *J. Mol. Biol.* **189**: 113-130.
- Studier, F. W., Rosenberg, A. H., Simon, M. N. y Dunn, J. J. 1979. Genetic and physical mapping in the early region of bacteriophage T7 DNA. *J. Mol. Biol.* **135**: 917-937.
- Stuvier, M.H., Bergsma, W.G., Arnberg, A.C., Van Amerongen, H., Van Grondelle, R. y Van der Vliet, P.C. 1992. Structural alterations of double-stranded DNA in complex with the adenovirus DNA binding protein. Implications for its function in DNA replication. *J. Mol. Biol.* **225**: 999-1.011.
- Swank, R.T. y Munkres, K.D. 1971. Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. *Anal. Biochem.* **39**: 462-477.
- Tabor, S. y Richardson, C.C. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **82**: 1074-1078.
- Tedin, K., Resch, A., Steiner, M. y Bläsi, U. 1995. Dual translational start motif evolutionarily conserved in the holin gene of *Bacillus subtilis* phage ϕ 29. *Virology* **206**: 479-484.
- Tiraby, J.G., Tiraby, E. y Fox, M. S. 1975. Pneumococcal bacteriophages. *Virology* **68**: 566-569.
- Tolun, A., Aleström, P. y Petterson, U. 1979. Sequenced of inverted terminal repetitions from different adenoviruses: demonstration of conserved sequences and homology between SA7 termini and SV40 DNA. *Cell* **17**: 705-713.
- Tomasz, A. 1967. Choline in the cell wall of a bacterium: novel type of polymer-linked choline in pneumococcus. *Science* **157**: 694-697.
- Tomasz, A. 1970. Requirement for protein synthesis during induction of competence. *J. Bacteriol.* **101**: 860-871.
- Tosi, M.E., Reilly, B.E. y Anderson, D. 1975. Morphogenesis of bacteriophage ϕ 29 of *Bacillus subtilis*: cleavage and assembly of the neck appendage protein. *J. Virol.* **16**: 1.282-1.295.
- Trautwetter, A., Ritzenthaler, P., Alatossava, T. y Mata-Gilsinger, M. 1986. Physical and genetic characterization of the genome of the *Lactobacillus lactis* bacteriophage LL-H. *J. Virol.* **59**: 551-555.

- Tsugita, A. e Inouye, M. 1968. Complete primary structure of phage lysozyme from *Escherichia coli* T4. *J. Mol. Biol.* **37**: 201-212.
- Twort, F. 1915. An investigation on the nature of the ultramicroscopic viruses. *Lancet* **189**: 1.241.
- Valpuesta, J.M., Donate, L., Mier, C., Herranz, L. y Carrascosa, J.L. 1993. RNA-mediated specificity of DNA packaging into hybrid $\lambda/\phi 29$ proheads. *EMBO J.* **12**: 4.453-4.459.
- Vasala, A., Dupont, L., Baumann, M., Ritzenthaler, P. y Alatossava, T. 1993. Molecular comparison of the structural proteins encoding gene clusters of two related *Lactobacillus delbrueckii* bacteriophages. *J. Virol.* **67**: 3.061-3.068.
- Verma, M. 1986. Molecular cloning and sequencing of lysozyme gene of bacteriophage SF6 of *Bacillus subtilis*. *Curr. Microbiol.* **13**: 299-301.
- Vieira, J. y Messing, J. 1982. The pUC plasmids a M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**: 259-268.
- Von Heijne, G. 1992. Membrane protein structure prediction hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**: 487-494.
- Waks, S. y Tomasz, A. 1978. Secretion of cell wall polymers into the growth medium of lysis-defective pneumococci during treatment with penicillin and other inhibitors of cell wall synthesis. *Antimicrob. Agents Chemother.* **13**: 293-301.
- Ward, L.J.H., Beresford, T.P.J., Lubbers, M.W., Jarvis, B.D.W. y Jarvis, A.W. 1993. Sequence analysis of the lysin gene region of the prolate lactococcal bacteriophage c2. *Can. J. Microbiol.* **39**: 767-774.
- Wichtwechkarn, J., Bailey, S., Bodley, J.W. y Anderson, D. 1989. Prohead RNA of bacteriophage $\phi 29$: size, stoichiometry and biological activity. *Nuc. Acids. Res.* **17**: 3459-3468.
- Winter, R.B. y Gold, L. 1983. Overproduction of bacteriophage Q β maturation (A₂) protein. *Cell* **33**: 877-885.
- Witte, A. y Lubitz, W. 1989. Biochemical characterization of the ϕ X174-protein-E-mediated lysis of the *Escherichia coli* Eur. *J. Biochem.* **180**: 393-398.
- Witte, A., Bläsi, U., Halfmann, G., Szostak, M., Wanner, G. y Lubitz, W. 1990. PhiX174 protein E-mediated lysis of *Escherichia coli*. *Biochimie* **72**: 191-200.
- Yamada, M., Fujisawa, H., Kato, H., Hamada, K. y Minagawa, T. 1986. Cloning and sequencing of the genetic right end of bacteriophage T3 DNA. *Virology* **151**: 350-361.
- Yanisch-Perron, C., Vieira, J. y Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.
- Yoo, S.K. e Ito, J. 1991a. Sequence requirements for protein-primed DNA replication of bacteriophage PRD1. *J. Mol. Biol.* **218**: 779-789.
- Yoo, S.K. e Ito, J. 1991b. Initiation of bacteriophage PRD1 DNA replication on single-stranded templates. *J. Mol. Biol.* **222**: 127-131.
- Young, R. 1992. Bacteriophage lysis: mechanism and regulation. *Microbiol. Rev.* **56**: 430-481.
- Young, R. y Bläsi, U. 1995. Holins: form and function in bacteriophage lysis. *FEMS Microbiol. Rev.* **17**: 191-205.
- Zaballos, A. y Salas, M. 1989. Functional domains in the bacteriophage $\phi 29$ terminal protein for interaction with the $\phi 29$ polymerase and with DNA. *Nucleic Acids Res.* **17**: 10.353-10.366.
- Zagotta, M.T. y Wilson, D.B. 1990. Oligomerization of the bacteriophage lambda S protein in the inner membrane of *E. coli*. *J. Bacteriol.* **172**: 912-921.
- Ziermann, R., Bartlett, B., Calendar, R. y Christie, G.E. 1994. Functions involved in bacteriophage P2-induced host cell lysis and identification of a new tail gen. *J. Bacteriol.* **176**: 4.974-4.984.
- Zink, R., Loessner, M.J. y Scherer S. 1995. Characterization of cryptic prophages (monocins) in *Listeria* and sequence analysis of a holin/endolysin gene. *Microbiol.* **141**: 2.577-2.584.
- Zuker, M. y Stiegler, P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**: 133-148.